THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXXI April, 1940 Number 4

COMMON AND DIFFERENTIAL FACTORS IN READING AND HEARING COMPREHENSION

ROBERT P. LARSEN AND D. D. FEDER

University of Illinois

The major emphasis of research in the psychology of reading has been upon the mechanical factors involved in the perception of visual stimuli. However, the process of reading consists of a great deal more than the mere apprehension of the printed symbols. Despite the fact that the meaning which a reader derives from the printed page depends upon the central process of comprehension, relatively little study has been made of the psychological factors in this process. Even less attention has been paid to the common and differential abilities employed in reading and hearing comprehension, since it has been assumed that once either visual or auditory stimuli are perceived, a central common process of comprehension comes into play. The present study was undertaken to determine whether or not certain psychological abilities differentiate between the processes involved in reading and hearing comprehension.

The subjects for the first part of this study were one hundred fifty-one university freshmen, chosen at random from the 1937–1938 first year speech students at the State University of Iowa.

The paragraph section of Form *B* of the *Nelson-Denny Reading Test* was used as the measure of reading comprehension (hereafter designated as NDRT). It consists of nine selections of approximately two hundred words, each composed of factual reading material on various topics. Each selection was read once, after which the subject answered four multiple-choice questions without referring again to the passage. Each succeeding selection increases by a known amount of difficulty, which permitted grouping of the items into three classes. The first three selections are relatively easy; the middle three selections are of moderate difficulty; and the last three selections are the most difficult. That differences among these three levels of difficulty are

all greater than zero is indicated by the critical ratios of three or more for the differences between any two means of the three classes.

The paragraph section of Form *A* of the *Nelson-Denny Reading Test* was used as the measure of hearing comprehension (hereafter designated as NDHT).* This test was comparable† with the reading comprehension test except that the selections were read by a trained speaker and recorded upon high fidelity phonograph records. The classifications of the levels of difficulty of Form *A* paralleled those of Form *B* of the reading test. Each of the selections in this test was heard once by the subjects after which they answered the questions referring to that selection.

Both the hearing comprehension and the reading comprehension tests were given to each subject during the same period. In order to give no preference to either test, the order of administration was rotated.

The reliability coefficient of the NDRT, based upon correlation of odd and even items was .82; that of the NDHT was .79.‡

A *Reading Memory Test*§ was used to measure memory for connected material in terms of ideas remembered. It consists of two separate selections each containing fifty ideas. A reliability coefficient of .69 was obtained for this test.

Each of the subjects read each selection once at his normal reading rate, immediately after which he wrote as much of it as possible using either his own words or those in the passage. The score was the number of ideas correctly remembered.

The *Hearing Memory Test* was comparable with the *Reading Memory Test*‖ except that each selection was read once to the subjects

* The terms "reading comprehension" and "hearing comprehension," whenever used hereafter, refer to scores received on the NDRT and the NDHT, respectively.

† The correlation between the two forms, when both were given as reading tests, was .81 ± .02. As an additional check, the two tests were given to thirty-eight randomly selected freshmen under the modified test conditions with the modes of presentation of the two forms reversed from those used in this study. The reading comprehension mean of 24.6 and the hearing comprehension mean of 23.6 achieved by this group are very comparable to the corresponding means achieved by the group used in the main study.

‡ The reliability coefficients referred to in this study were obtained by use of the Spearman-Brown prophecy formula.

§ Constructed by Robert P. Larsen.

‖ Both memory tests had previously been standardized upon a randomly selected group of freshmen.

by a trained speaker. The order of administering the two memory tests was rotated. A reliability coefficient of .67 was obtained.

The *Phonics Test*, constructed by Rogers,[5] was used to measure ability to pronounce artificial words. The test consists of one hundred items composed of an artificial word followed by a choice of five different pronunciations, only one of which was correct.

The *Complex Speech Sound Discrimination Test*, constructed by Hall,[4] was used as a measure of response to a complex auditory pattern. An artificial or coined word was presented correctly in a meaningful phrase which was read twice to the subject. Then the phrase was read six times, once correctly and once with the coined word containing each of five error types. The subject was instructed to decide which one of the six repetitions was correct, that is, determine which one was like the two original presentations.

The *Oral Reading Accuracy Test* was used in the present study because Swanson[6] showed that the inaccuracies made while reading orally are quite similar to those made while reading silently. It consists of three separate selections making a total of three hundred words. The three selections were of graded difficulty. The first two had been chosen and used by Swanson;[6] the third was a new one adapted especially for the present study.* The following oral reading errors were checked: Substitutions, repetitions, omissions, insertions, mispronunciations, and miscellaneous. This test was administered to each subject individually with instructions to read each selection aloud as though to himself and not to an audience.

All the tests were administered either individually or to small groups. The usual precautions were taken to prevent copying and to obtain strict obedience to directions. The usual time limits were not used, each subject being allowed to finish each test at his own normal rate.

All of the tests indicated in Table I show a slightly higher relationship to reading comprehension than to hearing comprehension. The inference part of the *Reading Comprehension Maturity Test E-2x*[2] showed a slightly higher relationship to hearing comprehension. The *Iowa Silent Reading Test* showed the highest relationship with both reading and hearing comprehension.

The corrected coefficient of .82 indicates a rather close relationship between reading and hearing scores. The mean number of questions

---

* The passage was selected and adapted by Robert P. Larsen.

answered correctly was twenty-five on the NDRT and twenty-three on the NDHT out of a possible thirty-six on each; the SD's were 5.3 and 4.4, respectively. The critical ratio of 5.4 of the difference between the NDRT mean and the NDHT mean indicates that there is a reliable tendency for the scores in reading to be higher than those in hearing for material of comparable difficulty.

TABLE I.—INTERCORRELATIONS OF READING COMPREHENSION AND HEARING COMPREHENSION WITH CERTAIN TESTS OF READING COMPREHENSION

Variables	1	2
2	.68 ± .03 .82 ± .03*	
3	.58 ± .04 .65 ± .03*	.53 ± .04 .62 ± .04*
4	.53 ± .04 .70 ± .03*	.50 ± .04 .67 ± .03*
5	.52 ± .04 .71 ± .03*	.50 ± .04 .69 ± .03*
6	.40 ± .05 .56 ± .04*	.42 ± .05 .60 ± .04*

$N = 151$

Variables.

1. Nelson-Denny reading test.
2. Nelson-Denny hearing test.
3. Iowa silent reading test.
4. Reading comprehension maturity, total.
5. Reading comprehension maturity, information.
6. Reading comprehension maturity, inference.

* Corrected for attenuation.

From the data in Table II it will be seen that the group low in scholastic aptitude was about equal in both reading comprehension and hearing comprehension. The median group showed some superiority in reading comprehension, but the difference was not statistically significant. The high group showed a superiority in reading comprehension. Thus it appears that scholastic aptitude varies directly with reading comprehension.

The data in Table III parallel those in Table II except that the group divisions were determined in terms of scores on the *Iowa Silent*

Reading Test. The trend is similar to that observed in Table II. There is a definite tendency for the better readers to comprehend relatively better when reading than when listening.

TABLE II.—EVALUATION OF THE DIFFERENCES BETWEEN MEAN READING COMPREHENSION AND MEAN HEARING COMPREHENSION SCORES FOR THE LOW, MEDIAN, AND HIGH GROUPS IN SCHOLASTIC APTITUDE*

Group	*N*	NDRT		NDHT		*CR*
		Mean	SD	Mean	SD	
Low	30	20.4 (57%)†	4.7	20.1 (56%)	4.3	.4
Median	30	25.6 (71%)	4.2	23.7 (66%)	3.2	2.6
High	30	29.0 (81%)	4.0	26.2 (73%)	4.6	4.3

* Scholastic aptitude is measured by the Iowa Qualifying Examinations which have been found consistently to yield correlation coefficients of .70 and better with first year achievement.[6]

† The number in parenthesis beneath each mean in this and subsequent tables represents that mean in terms of per cent of the total possible score.

TABLE III.—EVALUATION OF THE DIFFERENCES BETWEEN READING COMPREHENSION AND HEARING COMPREHENSION SCORES FOR THE LOW, MEDIAN, AND HIGH GROUPS IN READING ABILITY

Group	*N*	NDRT		NDHT		*CR*
		Mean	SD	Mean	SD	
Low	30	20.4 (57%)	4.3	20.0 (56%)	4.0	.8
Median	30	26.8 (74%)	4.1	24.0 (67%)	3.1	4.6
High	30	28.9 (80%)	4.4	25.8 (72%)	4.6	4.8

As previously noted, the items of the reading and hearing comprehension tests could be grouped into three classes of known difficulty. This grouping, reflected in Table IV, shows that the easy material may be comprehended as well when read as when heard. However, with material of moderate difficulty there is a tendency for reading compre-

hension to be superior; with difficult material, reading comprehension is markedly superior to hearing comprehension.

TABLE IV.—EVALUATION OF THE DIFFERENCES BETWEEN READING COMPREHENSION AND HEARING COMPREHENSION MEANS WHEN THE ITEMS ARE GROUPED ACCORDING TO LEVELS OF DIFFICULTY

Variable	*N*	NDRT		NDHT		*CR*
		Mean	SD	Mean	SD	
Total test	151	25 (70%)	5.3	23 (64%)	4.4	5.4
Easy material	151	9.8 (82%)	1.9	9.9 (83%)	1.7	.8
Moderate material	151	8.0 (67%)	2.4	7.4 (62%)	2.2	2.9
Difficult material	151	7.1 (59%)	2.4	5.8 (48%)	2.3	6.1

TABLE V.—EVALUATION OF THE DIFFERENCES BETWEEN READING COMPREHENSION AND HEARING COMPREHENSION MEANS FOR EASY AND DIFFICULT MATERIAL FOR THOSE LOW AND HIGH IN SCHOLASTIC APTITUDE

Group	*N*	Level	NDRT		NDHT		*CR*
			Mean	SD	Mean	SD	
Low	30	Easy	9.9 (81%)	1.8	10.2 (85%)	5.3	.6
	30	Difficult	5.6 (47%)	2.4	5.3 (44%)	2.0	.7
High	30	Easy	10.8 (90%)	1.3	10.4 (87%)	1.3	1.5
	30	Difficult	8.9 (74%)	1.8	7.0 (59%)	2.5	4.6

The data in Table V show that students high in scholastic aptitude comprehend difficult material better by reading than by hearing, and easy material almost equally well by either method. Students of low ability do slightly but not significantly better when listening to easy material, and are almost equal in comprehension of difficult material through either modality.

The group divisions in Table VI were made in terms of scores on the *Iowa Silent Reading Test.* The trend of the data is parallel to that observed in Table V.

TABLE VI.—EVALUATION OF THE DIFFERENCES BETWEEN READING COMPREHENSION AND HEARING COMPREHENSION FOR EASY AND DIFFICULT MATERIAL FOR THOSE LOW AND HIGH IN READING ABILITY

Group	*N*	Level	NDRT		NDHT		*CR*
			Mean	SD	Mean	SD	
Low	30	Easy	8.6 (72%)	2.2	8.6 (72%)	1.7	.1
	30	Difficult	5.6 (47%)	2.4	5.0 (42%)	1.7	1.7
High	30	Easy	10.5 (88%)	1.5	10.4 (87%)	1.4	.3
	30	Difficult	8.7 (73%)	2.0	6.6 (55%)	2.5	4.5

Nearly all of the subjects spent more time on each of the selections in the hearing test than on the corresponding selections in the reading test, thus giving a time superiority to hearing comprehension. It is logical to expect that if equal time had been allowed for both tests, an even greater superiority would have been found for reading comprehension.

In an attempt to find the relationship of sex differences to reading and hearing comprehension, all of the females ($N = 38$) were matched with an equal number of males in terms of reading scores. The males had a mean hearing score of 24 as compared with a mean hearing score of 22.6 for the females. The difference is not statistically significant ($CR = 1.6$).

On the assumption that the similarities and differences in psychological abilities involved in reading and hearing comprehension would best be illustrated in extreme cases, four groups of nine subjects each were selected as follows: (1) Poor readers-good hearers (PRGH's); (2) good readers-poor hearers (GRPH's); (3) poor readers-poor hearers (PRPH's); and (4) good readers-good hearers (GRGH's).

Each subject in the PRPH group was matched with one in the PRGH group in reading comprehension and with one in the GRPH group in hearing comprehension. Similarly each subject in the GRGH group was matched with a member of the GRPH group in reading comprehension and with one of the PRGH group in hearing comprehension.

All of the subjects in these groups took the tests indicated in Table VII.

TABLE VII.—MEAN PER CENT OF CORRECT ANSWERS ACHIEVED BY THE FOUR GROUPS ON ALL OF THE TESTS ADMINISTERED

Tests	PRPH	PRGH	GRPH	GRGH
NDRT	51	51	74	76
NDHT	45	73	43	73
Reading memory	43	52	56	58
Hearing memory	46	54	55	59
Phonics	43	50	51	55
Sound discrimination	82	87	85	86
Oral reading accuracy	92	92	95	95

The assumption that the GRGH's are superior to the PRPH's in both reading comprehension and hearing comprehension is justified by the high critical ratios of 11.4 and 12.1, respectively, obtained between their mean scores in comprehension. This superiority is reflected markedly in the scores in reading memory and phonics, but complex sound discrimination and oral reading accuracy show less differentiation.

Table VIII indicates that although the GRPH's show a very high superiority in reading comprehension and the PRGH's likewise reveal a high superiority in hearing comprehension, none of the abilities, with the exception of oral reading accuracy, seems to play an important rôle as a differential factor.

TABLE VIII.—EVALUATION OF THE MEAN PER CENT OF SUPERIORITY OF THE GRGH'S OVER THE PRPH'S AND THE GRPH'S OVER THE PRGH'S

Test	GRGH-PRPH		GRPH-PRGH	
	Difference	*CR*	Difference	*CR*
NDRT	25	11.4	23	10.3
NDHT	28	12.1	−30	−13.5
Reading memory	15	6.4	4	1.7
Hearing memory	13	5.6	1	.4
Phonics	10	4.3	1	.4
Sound discrimination	4	2.3	− 2	− 1.2
Oral reading accuracy	3	2.6	3	2.6

The data in Table IX corroborate the previous findings concerning the rôle of the psychological abilities in comprehension. Oral reading accuracy is the only measure which shows a statistically significant

difference between reading and hearing comprehension. Reading memory and phonics show the next highest relationship to reading comprehension. Hearing memory and complex sound discrimination show the highest positive relationship to hearing comprehension.

TABLE IX.—EVALUATION OF THE CORRELATIONS BETWEEN READING COMPREHENSION AND HEARING COMPREHENSION SCORES AND THE PSYCHOLOGICAL ABILITIES TESTED

Test	NDRT	NDHT	*CR*
Oral reading accuracy	.48 ± .13	−.18 ± .16	3.2
Reading memory	.42 ± .14	.14 ± .16	1.3
Phonics	.33 ± .15	.12 ± .16	.9
Hearing memory	.23 ± .16	.34 ± .15	.5
Sound discrimination	.18 ± .16	.16 ± .16	.1

The higher relationship between reading memory and reading comprehension, and between hearing memory and hearing comprehension, would point toward the validity of the memory tests as possible discriminative measures; however, the correlation between memory and comprehension is so low as to suggest that logical memory plays a small part in these comprehension tests. The almost identical correlations of reading comprehension and hearing comprehension with complex sound discrimination suggest that the latter test does not yield a satisfactory differentiation between the two methods of comprehension.

In order to find other possible factors that might be either common or differential to reading and hearing comprehension, all of the GRPH's and PRGH's were interviewed to ascertain their subjective reasons for the large differences between their reading and hearing comprehension scores.

Subjectively, the GRPH's reported two major comprehension difficulties: Three indicated inability to comprehend material heard, because the rate of speaking was not adjusted to individual needs; two mentioned inability to concentrate continuously upon material heard, although this difficulty was not prevalent while reading.

The following reasons for their low scores in reading comprehension were reported by the PRGH's: Four, one of whom had been out of school four years—inability to concentrate upon material read; two—intense dislike for reading; one—that he had developed a habit of skimming and very rapid reading while in high school; and one—that

in comprehending he depended to a great extent upon the speaker's intonations, pauses, etc., and since these were lacking in silent reading, this mode of comprehension was more difficult for him.

Among the PRPH's three reported inability to concentrate both while reading and listening; three admitted inability to concentrate only while reading; and one indicated difficulty only while listening.

Among the GRGH's two indicated difficulty in concentrating while reading, and one reported difficulty while listening.

The subjects were also asked to keep a record of the time spent in reading and listening (including lectures and conversation) for a typical school week. The GRPH's in one school day spent on the average five and one-half hours in reading, and four hours in listening; whereas, the mean of the PRGH's for one day was slightly less than four hours in reading and five hours in listening.

SUMMARY AND CONCLUSIONS

On the basis of the foregoing clinical-experimental investigation of the influence of certain psychological abilities upon reading comprehension and hearing comprehension, the following results were obtained:

(1) The correlation between the reading comprehension and hearing comprehension scores was $.62 \pm .04$ (corrected for attenuation $.82 \pm .02$). The reading comprehension raw score mean was 25 and the per cent of items answered correctly was seventy, while the hearing comprehension raw score mean was 23 and the per cent of items answered correctly was sixty-four. The difference between the raw score means was statistically significant.

(2) The superiority of performance in reading comprehension over that in hearing comprehension was found to be dependent upon the level of difficulty of the material. This superiority was slight for those low in scholastic aptitude and reading ability, but quite marked for those high in ability.

(3) Those low in scholastic aptitude and general reading ability comprehended almost as well by reading as by listening, whereas the median groups showed a slight superiority in favor of reading comprehension. The group high in scholastic aptitude showed a definite superiority in reading comprehension.

(4) Reading memory, hearing memory, complex sound discrimination, and phonics did not differentiate sufficiently to be considered significant conditioners of skill in reading and hearing comprehension.

Oral reading accuracy was the only differential factor showing a fair degree of relationship to reading comprehension, and a slight negative relationship to hearing comprehension.

(5) An interview with each of the thirty-six subjects revealed inability to concentrate as the most frequent cause of large differences in hearing and reading comprehension scores. This failure to concentrate played a part both as a common and as a differential factor in reading and listening. A time record revealed that those with relatively high scores in reading comprehension spent more time in reading, whereas those with high scores in hearing comprehension spent more time in listening.

On the basis of the data herein presented, certain tentative conclusions may be offered:

(1) This study supports the conclusions of previous studies that comprehension is largely a centrally-determined function operating independently of the mode of presentation of the material.

(2) The data indicate the following implications and suggestions for remedial training: (*a*) Reading seems to be largely language or thought activity; (*b*) the oral reading accuracy tests may be of some aid in the diagnosis of reading difficulties; (*c*) phonics, complex sound discrimination, hearing memory, and reading memory play a small part in hearing or reading comprehension difficulties; (*d*) students with low scholastic aptitude or reading scores should be taught to adapt their reading rate to the difficulty level of the material; (*e*) certain students should be given remedial training in listening.

(3) Material presented orally should be somewhat less difficult than visually presented material in order to be comprehended with equal effectiveness.

It appears from this investigation that none of the abilities studied is alone capable of accounting for the differences observed in reading and hearing comprehension. Nevertheless, such differences have been shown to exist in significant degree. The future course of effective teaching and remedial work in reading must undoubtedly be influenced by knowledge of the psychological elements of which the comprehension skills are composed. Therefore, further research is necessary concerning other psychological abilities involved in comprehension. The importance of the language factor in comprehension as revealed by this and other studies further suggests that little improvement in the central comprehension process can be obtained by excessive mechanization of the reading skills.

BIBLIOGRAPHY

1. Anderson, I. H. and Fairbanks, G.: "Common and differential factors in reading vocabulary and hearing vocabulary." *Journal of Educational Research*, Vol. XXX, 1937, No. 5, pp. 317–325.
2. Feder, D. D.: "Comprehension Maturity Tests—A new technique in mental measurement." *Journal of Educational Psychology*, Vol. XXIX, November, 1938, pp. 597–606.
3. Feder, D. D.: *The use of Iowa Placement Examinations in the prediction of success.* (Pamphlet) University of Iowa, Bureau of Educational Research and Service.
4. Hall, M. E.: *Auditory factors in functional articulation speech defects.* Unpublished Ph. D. thesis, State University of Iowa, 1938.
5. Rogers, M. V.: *Phonics ability and its relationship to certain aspects of reading at the college level.* Unpublished Ph. D. thesis, State University of Iowa, 1937.
6. Swanson, D. E.: "Common elements in silent and oral reading." *Psychological Monographs*, Vol. XLVIII, 1937, No. 3, pp. 36–60.

DEGREES OF FREEDOM

HELEN M. WALKER

Associate Professor of Education, Teachers College, Columbia University

A concept of central importance to modern statistical theory which few textbooks have attempted to clarify is that of "degrees of freedom." For the mathematician who reads the original papers in which statistical theory is now making such rapid advances, the concept is a familiar one needing no particular explanation. For the person who is unfamiliar with N-dimensional geometry or who knows the contributions to modern sampling theory only from secondhand sources such as textbooks, this concept often seems almost mystical, with no practical meaning.

Tippett, one of the few textbook writers who attempt to make any general explanation of the concept, begins his account (p. 64) with the sentence, "This conception of *degrees of freedom* is not altogether easy to attain, and we cannot attempt a full justification of it here; but we shall show its reasonableness and shall illustrate it, hoping that as a result of familiarity with its use the reader will appreciate it." Not only do most texts omit all mention of the concept but many actually give incorrect formulas and procedures because of ignoring it.

In the work of modern statisticians, the concept of degrees of freedom is not found before "Student's" paper of 1908, it was first made explicit by the writings of R. A. Fisher, beginning with his paper of 1915 on the distribution of the correlation coefficient, and has only within the decade or so received general recognition. Nevertheless the concept was familiar to Gauss and his astronomical associates. In his classical work on the *Theory of the Combination of Observations* (Theoria Combinationis Observationum Erroribus Minimis Obnoxiae) and also in a work generalizing the theory of least squares with reference to the combination of observations (Ergänzung zur Theorie der den kleinsten Fehlern unterworfen Combination der Beobachtungen, 1826), he states both in words and by formula that the number of observations is to be decreased by the number of unknowns estimated from the data to serve as divisor in estimating the standard error of a set of observations, or in our terminology $\sigma^2 = \frac{\Sigma x^2}{N - r}$ where r is the number of parameters to be estimated from the data.

The present paper is an attempt to bridge the gap between mathematical theory and common practice, to state as simply as possible

what degrees of freedom represent, why the concept is important, and how the appropriate number may be readily determined. The treatment has been made as non-technical as possible, but this is a case where the mathematical notion is simpler than any non-mathematical interpretation of it. The paper will be developed in four sections: (I) The freedom of movement of a point in space when subject to certain limiting conditions, (II) The representation of a statistical sample by a single point in *N*-dimensional space, (III) The import of the concept of degrees of freedom, and (IV) Illustrations of how to determine the number of degrees of freedom appropriate for use in certain common situations.

I. THE FREEDOM OF MOVEMENT OF A POINT IN SPACE WHEN SUBJECT TO CERTAIN LIMITING CONDITIONS

As a preliminary introduction to the idea, it may be helpful to consider the freedom of motion possessed by certain familiar objects, each of which is treated as if it were a mere moving point without size. A drop of oil sliding along a coil spring or a bead on a wire has only one degree of freedom for it can move only on a one-dimensional path, no matter how complicated the shape of that path may be. A drop of mercury on a plane surface has two degrees of freedom, moving freely on a two-dimensional surface. A mosquito moving freely in three-dimensional space, has three degrees of freedom.

Considered as a moving point, a railroad train moves backward and forward on a linear path which is a one-dimensional space lying on a two-dimensional space, the earth's surface, which in turn lies within a three-dimensional universe. A single coördinate, distance from some origin, is sufficient to locate the train at any given moment of time. If we consider a four-dimensional universe in which one dimension is of time and the other three dimensions of space, two coördinates will be needed to locate the train, distance in linear units from a spatial origin and distance in time units from a time origin. The train's path which had only one dimension in a space universe has two dimensions in a space-time universe.

A canoe or an automobile moves over a two-dimensional surface which lies upon a three-dimensional space, is a section of a three-dimensional space. At any given moment, the position of the canoe, or auto, can be given by two coördinates. Referred to a four-dimensional space-time universe, three coördinates would be needed to give

its location, and its path would be a space of three dimensions, lying upon one of four.

In the same sense an airplane has three degrees of freedom in the usual universe of space, and can be located only if three coördinates are known. These might be latitude, longitude, and altitude; or might be altitude, horizontal distance from some origin, and an angle; or might be direct distance from some origin, and two direction angles. If we consider a given instant of time as a section through the space-time universe, the airplane moves in a four-dimensional path and can be located by four coördinates, the three previously named and a time coördinate.

The degrees of freedom we have been considering relate to the motion of a point, or freedom of translation. In mechanics freedom of *rotation* would be equally important. A point, which has position only, and no size, can be translated but not rotated. A real canoe can turn over, a real airplane can turn on its axis or make a nose dive, and so these real bodies have degrees of freedom of rotation as well as of translation. The parallelism between the sampling problems we are about to discuss and the movement of bodies in space can be brought out more clearly by discussing freedom of translation, and disregarding freedom of rotation, and that has been done in what follows.

If you are asked to choose a pair of numbers (x, y) at random, you have complete freedom of choice with regard to each of the two numbers, have two degrees of freedom. The number pair may be represented by the coördinates of a point located in the x, y plane, which is a two-dimensional space. The point is free to move anywhere in the horizontal direction parallel to the xx' axis, and is also free to move anywhere in the vertical direction, parallel to the yy' axis. There are two independent variables and the point has two degrees of freedom.

Now suppose you are asked to choose a pair of numbers whose sum is 7. It is readily apparent that only one number can be chosen freely, the second being fixed as soon as the first is chosen. Although there are two variables in the situation, there is only one independent variable. The number of degrees of freedom is reduced from two to one by the imposition of the condition $x + y = 7$. The point is not now free to move anywhere in the xy plane but is constrained to remain on the line whose graph is $x + y = 7$, and this line is a one-dimensional space lying in the original two-dimensional space.

Suppose you are asked to choose a pair of numbers such that the sum of their squares is 25. Again it is apparent that only one number can be chosen arbitrarily, the second being fixed as soon as the first is chosen. The point represented by a pair of numbers must lie on a circle with center at the origin and radius 5. This circle is a one-dimensional space lying in the original two-dimensional plane. The point can move only forward or backward along this circle, and has one degree of freedom only. There were two numbers to be chosen ($N = 2$) subject to one limiting relationship ($r = 1$) and the resultant number of degrees of freedom is $N - r = 2 - 1 = 1$.

Suppose we simultaneously impose the two conditions $x + y = 7$ and $x^2 + y^2 = 25$. If we solve these equations algebraically we get only two possible solutions, $x = 3$, $y = 4$, or $x = 4$, $y = 3$. Neither variable can be chosen at will. The point, once free to move in two directions, is now constrained by the equation $x + y = 7$ to move only along a straight line, and is constrained by the equation $x^2 + y^2 = 25$ to move only along the circumference of a circle, and by the two together is confined to the intersection of that line and circle. There is no freedom of motion for the point. $N = 2$ and $r = 2$. The number of degrees of freedom is $N - r = 2 - 2 = 0$.

Consider now a point (x, y, z) in three-dimensional space ($N = 3$). If no restrictions are placed on its coördinates, it can move with freedom in each of three directions, has three degrees of freedom. All three variables are independent. If we set up the restriction $x + y + z = c$, where c is any constant, only two of the numbers can be freely chosen, only two are independent observations. For example, let $x - y - z = 10$. If now we choose, say, $x = 7$ and $y = 9$, then z is forced to be -12. The equation $x - y - z = c$ is the equation of a plane, a two-dimensional space cutting across the original three-dimensional space, and a point lying on this space has two degrees of freedom. ($N - r = 3 - 1 = 2$.) If the coördinates of the (x, y, z) point are made to conform to the condition $x^2 + y^2 + z^2 = k$, the point will be forced to lie on the surface of a sphere whose center is at the origin and whose radius is $\sqrt{k}$. The surface of a sphere is a two-dimensional space. ($N = 3$, $r = 1$, $N - r = 3 - 1 = 2$.)

If both conditions are imposed simultaneouly, the point can lie only on the intersection of the sphere and the plane, that is, it can move only along the circumference of a circle, which is a one-dimensional figure lying in the original space of three dimensions. ($N - r = 3 - 2 = 1$.) Considered algebraically, we note that solving the

pair of equations in three variables leaves us a single equation in two variables. There can be complete freedom of choice for one of these, no freedom for the other. There is one degree of freedom.

The condition $x = y = z$ is really a pair of independent conditions, $x = y$ and $x = z$, the condition $y = z$ being derived from the other two. Each of these is the equation of a plane, and their intersection gives a straight line through the origin making equal angles with the three axes. If $x = y = z$, it is clear that only one variable can be chosen arbitrarily, there is only one independent variable, the point is constrained to move along a single line, there is one degree of freedom.

These ideas must be generalized for N larger than 3, and this generalization is necessarily abstract. Too ardent an attempt to visualize the outcome leads only to confusion. Any set of N numbers determine a single point in N-dimensional space, each number providing one of the N coördinates of that point. If no relationship is imposed upon these numbers, each is free to vary independently of the others, and the number of degrees of freedom is N. Every necessary relationship imposed upon them reduces the number of degrees of freedom by one. Any equation of the first degree connecting the N variables is the equation of what may be called a hyperplane (Better not try to visualize it!) and is a space of $N - 1$ dimensions. If, for example, we consider only points such that the sum of their coördinates is constant, $\Sigma X = c$, we have limited the point to an $N - 1$ space. If we consider only points such that $\Sigma(X - M)^2 = k$, the locus is the surface of a hypersphere with center at the origin and radius equal to $\sqrt{k}$. This surface is called the locus of the point and is a space of $N - 1$ dimensions lying within the original N space. If r such conditions should be imposed simultaneously, the point would be confined to the intersection of the various loci, which would be a space of $N - r$ dimensions lying within the original N space. The number of degrees of freedom would be $N - r$.

II. THE REPRESENTATION OF A STATISTICAL SAMPLE BY A POINT IN N-DIMENSIONAL SPACE

If any N numbers can be represented by a single point in a space of N dimensions, obviously a statistical sample of N cases can be so represented by a single sample point. This device, first employed by R. A. Fisher in 1915 in a celebrated paper ("Frequency distribution of the values of the correlation coefficient in samples from an indefi-

nitely large population") has been an enormously fruitful one, and must be understood by those who hope to follow recent developments.

Let us consider a sample space of N dimensions, with the origin taken at the true population mean, which we will call μ, so that $X_1 - \mu = x_1$, $X_2 - \mu = x_2$, etc., where $X_1, X_2, \ldots X_N$ are the raw scores of the N individuals in the sample. Let M be the mean and s the

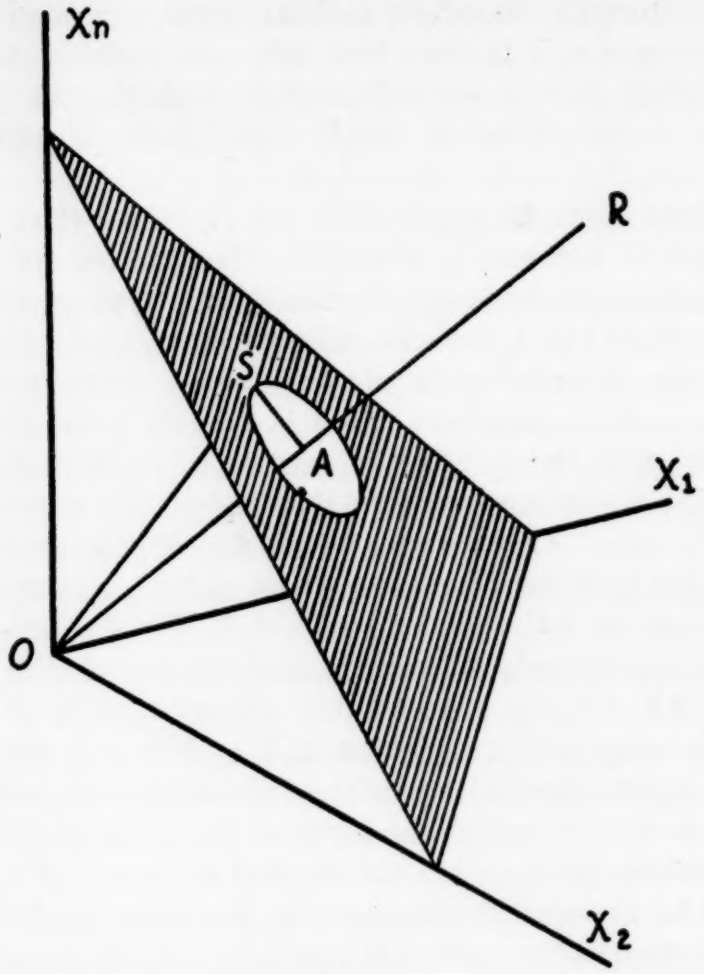

FIG. 1.

standard deviation of a sample of N cases. Any set of N observations determines a single sample point, such as S. This point has N degrees of freedom if no conditions are imposed upon its coördinates.

All samples with the same mean will be represented by sample points lying on the hyper-plane $(X_1 - \mu) + (X_2 - \mu) + \cdots + (X_N - \mu) = N(M - \mu)$, or $\Sigma X = NM$, a space of $N - 1$ dimensions.

If all cases in a sample were exactly uniform, the sample point would lie upon the line $X_1 - \mu = X_2 - \mu = X_3 - \mu = \cdots = X_N - \mu = M - \mu$ which is the line OR in Fig. 1, a line making equal

angles with all the coördinate axes. This line cuts the plane $\Sigma X = NM$ at right angles at a point we may call A. Therefore, A is a point whose coördinates are each equal to $M - \mu$. By a well-known geometric relationship,

$$\overline{OS}^2 = (X_1 - \mu)^2 + (X_2 - \mu)^2 + \cdots + (X_N - \mu)^2$$
$$\overline{OA}^2 = N(M - \mu)^2$$
$$\overline{OS}^2 = \overline{OA}^2 + \overline{AS}^2$$
$$\overline{AS}^2 = \Sigma(X - \mu)^2 - N(M - \mu)^2 = \Sigma X^2 - NM^2 = Ns^2$$

Therefore, $OA = (M - \mu)\sqrt{N}$ and $AS = s\sqrt{N}$.

The ratio $\frac{OA}{AS}$ is thus $\frac{M - \mu}{s}$ and is proportional to the ratio of the amount by which a sample mean deviates from the population mean to its own standard error. The fluctuation of this ratio from sample to sample produces what is known as the t-distribution.

For computing the variability of the scores in a sample around a population mean which is known *a priori*, there are available N degrees of freedom because the point S moves in N-dimensional space about O; but for computing the variability of these same scores about the mean of their own sample, there are available only $N - 1$ degrees of freedom, because one degree has been expended in the computation of that mean, so that the point S moves about A in a space of only $N - 1$ dimensions.

Fisher has used these spatial concepts to derive the sampling distribution of the correlation coefficient. The full derivation is outside the scope of this paper but certain aspects are of interest here. When we have N individuals each measured in two traits, it is customary to represent the N pairs of numbers by a correlation diagram of N points in two-dimensional space. The same data can, however, be represented by two points in N-dimensional space, one point representing the N values of X and the other the N values of Y. In this frame of reference the correlation coefficient can be shown to be equal to the cosine of the angle between the vectors to the two points, and to have $N - 2$ degrees of freedom.

III. THE IMPORT OF THE CONCEPT

If the normal curve adequately described all sampling distributions, as some elementary treatises seem to imply, the concept of degrees of freedom would be relatively unimportant, for this number does not

appear in the equation of the normal curve, the shape of the curve being the same no matter what the size of the sample. In certain other important sampling distributions—as for example the Poisson—the same thing is true, that the shape of the distribution is independent of the number of degrees of freedom involved. Modern statistical analysis, however, makes much use of several very important sampling distributions for which the shape of the curve changes with the effective size of the sample. In the equations of such curves, the number of degrees of freedom appears as a parameter (called n in the equations which follow) and probability tables built from these curves must be entered with the correct value of n. If a mistake is made in determining n from the data, the wrong probability value will be obtained from the table, and the significance of the test employed will be wrongly interpreted. The Chi-square distribution, the t-distribution, and the F and z distributions are now commonly used even in elementary work, and the table for each of these must be entered with the appropriate value of n.

Let us now look at a few of these equations to see the rôle played in them by the number of degrees of freedom. In the formulas which follow, C represents a constant whose value is determined in such a way as to make the total area under the curve equal to unity. Although this constant involves the number of degrees of freedom, it does not need to be considered in reading probability tables because, being a constant multiplier, it does not affect the proportion of area under any given segment of the curve, but serves only to change the scale of the entire figure.

Normal Curve.

$$y = C_1 e^{-\frac{x^2}{2\sigma^2}}$$

The number of degrees of freedom does not appear in the equation, and so the shape of the curve is independent of it. The only variables to be shown in a probability table are x/σ and y or some function of y such as a probability value.

Chi-square.

$$y = C_2(\chi^2)^{\frac{n-2}{2}} e^{-\frac{\chi^2}{2}}$$

The number of degrees of freedom appears in the exponent. When $n = 1$, the curve is J-shaped. When $n = 2$, the equation reduces to

$y = C_2 e^{-\frac{\chi^2}{2}}$ and has the form of the positive half of a normal curve. The curve is always positively skewed, but as n increases it becomes more and more nearly like the normal, and becomes approximately normal when n is 30 or so. A probability table must take account of three variables, the size of Chi-square, the number of degrees of freedom, and the related probability value.

t-distribution.

$$y = C_3\left(1 + \frac{t^2}{n}\right)^{-\frac{(n+1)}{2}}$$

The number of degrees of freedom appears both in the exponent and in the fraction t^2/n. The curve is always symmetrical, but is more peaked than the normal when n is small. This curve also approaches the normal form as n increases. A table of probability values must be entered with the computed value of t and also with the appropriate value of n. A few selected values will show the comparison between estimates of significance read from a table of the normal curve and a t-table.

For a normal curve, the proportion of area in both tails of the curve beyond 3σ is .0027. For a t-distribution the proportion is as follows:

n	1	2	5	10	20
p	.204	.096	.030	.014	.007

Again, for a normal curve, the point such that .01 of the area is in the tails, is 2.56σ from the mean.

For a t-distribution, the position of this point is as follows:

n	1	2	3	5	10	20	30
x/σ	63.6	9.9	5.8	4.0	3.2	2.8	2.75

F-distribution and z-distribution.

$$y = C_4 \frac{F^{\frac{n_1-2}{2}}}{(n_1F + n_2)^{\frac{n_1+n_2}{2}}} \quad \text{and} \quad y = C_5 \frac{e^{n_1 z}}{(n_1 e^{2z} + n_2)^{\frac{n_1+n_2}{2}}}$$

In each of these equations, which provide the tables used in analysis of variance problems, there occurs not only the computed value of F

(or of z), but also the two parameters n_1 and n_2, n_1 being the number of degrees of freedom for the mean square in the numerator of F and n_2 the number of degrees of freedom for that in the denominator. Because a probability table must be entered with all three, such a table often shows the values for selected probability values only. The tables published by Fisher give values for $p = .05$, $p = .01$, and $p = .001$; those by Snedecor give $p = .05$ and $p = .01$.

Sampling Distribution of r.

This is a complicated equation involving as parameters the true correlation in the population, ρ; the observed correlation in the sample, r; and the number of degrees of freedom. If $\rho = 0$ the distribution is symmetrical. If $\rho = 0$ and n is large, the distribution becomes normal. If $\rho \neq 0$ and n is small the curve is definitely skewed. David's *Tables of the Correlation Coefficient* (Issued by the Biometrika Office, University College, London, 1938) must be entered with all three parameters.

IV. DETERMINING THE APPROPRIATE NUMBER OF DEGREES OF FREEDOM IN CERTAIN TYPICAL PROBLEMS

A universal rule holds: The number of degrees of freedom is always equal to the number of observations minus the number of necessary relations obtaining among these observations. In geometric terms, the number of observations is the dimensionality of the original space and each relationship represents a section through that space restricting the sample point to a space of one lower dimension. Imposing a relationship upon the observations is equivalent to estimating a parameter from them. For example, the relationship $\Sigma X = NM$ indicates that the mean of the population has been estimated from observations. The number of degrees of freedom is also equal to the number of independent observations, which is the number of original observations minus the number of parameters estimated from them.

Standard Error of a Mean.—This is $\sigma_{mean} = \sigma/\sqrt{N}$ when σ is known for the population. As σ is seldom known *a priori*, we are usually forced to make use of the observed standard deviation in the sample, which we will call s. In this case $\sigma_{mean} \doteq s/\sqrt{N-1}$, one degree of freedom being lost because deviations have been taken around the sample mean, so that we have imposed one limiting relationship, $\Sigma X = NM$, and have thus restricted the sample point to a hyperplane of $N - 1$ dimensions.

Without any reference to geometry, it can be shown by an algebraic solution that $s\sqrt{N} \doteq \sigma\sqrt{N-1}$. (The symbol $\doteq$ is to be read "tends to equal" or "approximates.")

Goodness of Fit of Normal Curve to a Set of Data.—The number of observations is the number of intervals in the frequency distribution for which an observed frequency is compared with the frequency to be expected on the assumption of a normal distribution. If this normal curve has an arbitrary mean and standard deviation agreed upon in advance, the number of degrees of freedom with which we enter the Chi-square table to test goodness of fit is one less than the number of intervals. In this case one restriction is imposed; namely $\Sigma f = \Sigma f'$, where f is an observed and f' a theoretical frequency. If, however, as is more common, the theoretical curve is made to conform to the observed data in its mean and standard deviation, two additional restrictions are imposed; namely $\Sigma fX = \Sigma f'X$ and $\Sigma f(X - M)^2 = \Sigma f'(X - M)^2$, so that the number of degrees of freedom is three less than the number of intervals compared. It is clear that when the curves are made to agree in mean and standard deviation, the discrepancy between observed and theoretical frequencies will be reduced, so the number of degrees of freedom in relation to which that discrepancy is interpreted should also be reduced.

Relationship in a Contingency Table.—Suppose we wish to test the existence of a relationship between trait A, for which there are three categories, and trait B, for which there are five, as shown in Fig. 2. We have fifteen cells in the table, giving us fifteen observations, inasmuch as an "observation" is now the frequency in a single cell. If we want to ask whether there is sufficient evidence to believe that in the population from which this sample is drawn A and B are independent, we need to know the cell frequencies which would be expected under that hypothesis. There are then fifteen comparisons to be made between observed frequencies and expected frequencies. But are all fifteen of these comparisons independent?

If we had *a priori* information as to how the traits would be distributed theoretically, then all but one of the cell comparisons would be independent, the last cell frequency being fixed in order to make up the proper total of one hundred fifty, and the degrees of freedom would be $15 - 1 = 14$. This is the situation Karl Pearson had in mind when he first developed his Chi-square test of goodness of fit, and Table XII in Vol. I of his *Tables for Statisticians and Biometri-*

cians is made up on the assumption that the number of degrees of freedom is one less than the number of observations. To use it when that is not the case we merely readjust the value of n with which we enter the table.

In practice we almost never have *a priori* estimates of theoretical frequencies, but must obtain them from the observations themselves, thus imposing restrictions on the number of independent observations and reducing the degrees of freedom available for estimating reliability. In this case, if we estimate the theoretical frequencies from the data, we would estimate the frequency $f'_{11} = (20)(40)/150$ and others in similar fashion. Getting the expected cell frequencies from the observed marginal frequencies imposes the following relationships:

$$
\begin{aligned}
(a)\ & f_{11} + f_{21} + f_{31} + f_{41} + f_{51} = 40 \\
& f_{12} + f_{22} + f_{32} + f_{42} + f_{52} = 60 \\
& f_{13} + f_{23} + f_{33} + f_{43} + f_{53} = 50 \\
(b)\ & f_{11} + f_{12} + f_{13} = 20 \\
& f_{21} + f_{22} + f_{23} = 20 \\
& f_{31} + f_{32} + f_{33} = 35 \\
& f_{41} + f_{42} + f_{43} = 30 \\
& f_{51} + f_{52} + f_{53} = 50 \\
(c)\ & f_{11} + f_{21} + \cdots + f_{51} + f_{12} + \cdots + f_{53} = 150
\end{aligned}
$$

	A_1	A_2	A_3	
B_1	12	3	5	20
B_2	3	6	11	20
B_3	3	30	2	35
B_4	9	14	7	30
B_5	13	7	25	45
	40	60	50	150

FIG. 2.—Observed joint frequency distribution of two traits *A* and *B*.

	A_1	A_2	A_3	
B_1	f_{11}	f_{12}	f_{13}	20
B_2	f_{21}	f_{22}	f_{23}	20
B_3	f_{31}	f_{32}	f_{33}	35
B_4	f_{41}	f_{42}	f_{43}	30
B_5	f_{51}	f_{52}	f_{53}	45
	40	60	50	150

FIG. 3.—Observed marginal frequencies of two traits *A* and *B*.

At first sight, there seem to be nine relationships, but it is immediately apparent that (*c*) is not a new one, for it can be obtained either by adding the three (*a*) equations or the five (*b*) equations. Also any

one of the remaining eight can be obtained by appropriate manipulation of the other seven. There are then only seven independent necessary relationships imposed upon the cell frequencies by requiring them to add up to the observed marginal totals. Thus $n = 15 - 7 = 8$, and if we compute Chi-square, we must enter the Chi-square table with eight degrees of freedom. The same result can be obtained by noting that two entries in each row and four in each column can be chosen arbitrarily and there is then no freedom of choice for the remaining entries.

In general in a contingency table, if c = number of columns and r = number of rows, the number of degrees of freedom is $n = (c - 1)(r - 1)$ or $n = rc - (r + c - 1)$.

Variance in a Correlation Table.—Suppose we have a scatter diagram with c columns, the frequencies in the various columns being $n_1, n_2, \ldots n_c$, the mean values of Y for the columns being $m_1, m_2, \ldots m_c$, and the regression values of Y estimated from X being $\tilde{Y}_1, \tilde{Y}_2, \ldots \tilde{Y}_c$. Thus for any given column, the sum of the Y's is $\sum_1^{n_i} fY = n_i m_i$. For the entire table $N = n_1 + n_2 + \cdots + n_c$, $NM = \sum_1^{c} \sum_1^{n_i} fY$, so that $NM = n_1 m_1 + n_2 m_2 + \cdots + n_c m_c$.

Now we may be interested in the variance of all the scores about the total mean, of all the scores about their own column means, of all the scores about the regression line, of regressed values about the total mean, of column means about the total mean, or of column means about the regression line, and we may be interested in comparing two such variances. It is necessary to know how many degrees of freedom are available for such comparisons.

(*a*) *Total Variance.*—For the variance of all scores about the total mean, this is $s^2 = \frac{1}{N}\sum_1^N (Y - M)^2$, we have N observations and only one restriction; namely, $\Sigma fY = NM$. Thus there are $N - 1$ degrees of freedom.

(*b*) *Variance of Regressed Values about Total Mean.*—The equation for the regressed values being $\tilde{Y} - M_y = r\frac{s_x}{s_y}(X - M_x)$, it is clear that as soon as x is known, y is also known. The sample point can move only on a straight line. There is only one degree of freedom available for the variance of regressed values.

(*c*) *Variance of Scores about Regression Line.*—There are N residuals of the form $Y - \tilde{Y}$ and their variance is the square of the standard error of estimate, or $s^2_y(1 - r^2_{xy})$. There are N observations and two restrictions; namely,

$$\Sigma f(Y - \tilde{Y}) = 0$$

and

$$\Sigma f(Y - \tilde{Y})^2 = Ns^2_y(1 - r^2_{xy}).$$

Thus there are $N - 2$ degrees of freedom available.

(*d*) *Variance of Scores about Column Means.*—If from each score we subtract not the regression value but the mean of the column in which it stands, the variance of the residuals thus obtained will be $s_y^2(1 - E^2)$ where E is the correlation ratio obtained from the sample. There are N such residuals. For each column we have the restriction $\sum_1^{n_i} fY = n_i m_i$ making c restrictions in all. The number of degrees of freedom for the variance within columns is therefore $N - c$.

(*e*) *Variance of Column Means about Total Means.*—To compute this variance we have c observations, *i.e.*, the means of c columns, restricted by the single relation $NM = \sum_1^c n_i m_i$, and therefore have $c - 1$ degrees of freedom. The variance itself can be proved to be $s_y^2 E^2$, and represents the variance among the means of columns.

(*f*) *Variance of Column Means about Regression Line.*—If for each column we find the difference $m_i - \tilde{Y}_i$ between the column mean and the regression value, and then find $\frac{1}{N}\sum_1^c f_i(m_i - \tilde{Y}_i)^2$, the result will be $s_y^2(E^2 - r^2)$ which is a variance representing the departure of the means from linearity. There is one such difference for each column, giving us c observations, and these observations are restricted by the two relationships, $\sum_1^c f_i(m_i - \tilde{Y}_i) = 0$

and

$$\sum_1^c f_i(m_i - \tilde{Y}_i)^2 = Ns_y^2(E^2 - r^2).$$

Therefore, we have $c - 2$ degrees of freedom.

The following scheme shows these relationships in summary form:

Source of variation	Formula	Degrees of freedom
(*d*) Scores about column means	$s^2(1 - E^2)$	$N - c$
(*e*) Means about total mean	s^2E^2	$c - 1$
(*a*) Total	s^2	$N - 1$
(*c*) Scores about regression line	$s^2(1 - r^2)$	$N - 2$
(*b*) Regressed values about total mean	s^2r^2	1
(*a*) Total	s^2	$N - 1$
(*d*) Scores about column means	$s^2(1 - E^2)$	$N - c$
(*f*) Column means about regression line	$s^2(E^2 - r^2)$	$c - 2$
(*c*) Scores about regression line	$s^2(1 - r^2)$	$N - 2$
(*b*) Regressed values about total mean	s^2r^2	1
(*f*) Column means about regression line	$s^2(E^2 - r^2)$	$c - 2$
(*e*) Column means about total mean	s^2E^2	$c - 1$
(*b*) Regressed values about total mean	s^2r^2	1
(*f*) Column means about regression line	$s^2(E^2 - r^2)$	$c - 2$
(*d*) Scores about column means	$s^2(1 - E^2)$	$N - c$
(*a*) Total	s^2	$N - 1$

It is apparent that these variances have additive relationships and that their respective degrees of freedom have exactly the same additive relationships.

Tests Based on Ratio of Two Variances.—From any pair of these additive variances, we may make an important statistical test. Thus, to test whether linear correlation exists in the population or not, we may divide $\frac{s^2r^2}{1}$ by $\frac{s^2(1 - r)^2}{N - 2}$ obtaining $\frac{r^2(N - 2)}{1 - r^2}$. To test whether a relationship measureable by the correlation ratio exists in the population, we may divide $\frac{s^2E^2}{c - 1}$ by $\frac{s^2(1 - E^2)}{N - c}$ obtaining $\frac{E^2}{1 - E^2} \cdot \frac{N - c}{c - 1}$. To test whether correlation is linear, we may divide $\frac{s^2(E^2 - r^2)}{c - 2}$ by

$\frac{s^2r^2}{1}$ obtaining $\frac{E^2 - r^2}{r^2(c-2)}$ or may divide $\frac{s^2(E^2 - r^2)}{c-2}$ by $\frac{s^2(1-E^2)}{N-c}$ obtaining $\frac{E^2 - r^2}{1 - E^2} \cdot \frac{N-c}{c-2}$. In each case, the resulting value is referred to Snedecor's F-table which must be entered with the appropriate number of degrees of freedom for each variance. Or we may find the logarithm of the ratio to the base e, take half of it, and refer the result to Fisher's z-table, which also must be entered with the appropriate number of degrees of freedom for each variance.

Partial Correlation.—For a coefficient of correlation of zero order, there are $N - 2$ degrees of freedom. This is obvious, since a straight regression line can be fitted to any two points without residuals, and the first two observations furnish no estimate of the size of r. For each variable that is held constant in a partial correlation, one additional degree of freedom is lost, so that for a correlation coefficient of the pth order, the degrees of freedom are $N - p - 2$. This places a limit upon the number of meaningful interrelationships which can be obtained from a small sample. As an extreme illustration, suppose twenty-five variables have been measured for a sample of twenty-five cases only, and all the intercorrelations computed, as well as all possible partial correlations—the partials of the twenty-third order will of necessity be either $+1$ or -1, and thus are meaningless. Each such partial will be associated with $25 - 23 - 2$ degrees of freedom. If the partial were not $+1$ or -1 the error variance $\frac{\sigma^2(1 - r^2)}{N - p - 2}$ would become infinite, a fantastic situation.

BIBLIOGRAPHY

Dawson, S.: *An Introduction to the Computation of Statistics.* University of London Press, 1933, p. 114. No general discussion. Gives rule for χ^2 only.

Ezekiel, M.: *Methods of Correlation Analysis.* John Wiley & Sons, 1930, p. 121.

Fisher, R. A.: "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population." *Biometrika,* Vol. x, 1915, pp. 507–521. First application of n-dimensional geometry to sampling theory.

Fisher, R. A.: *Statistical Methods for Research Workers.* Oliver and Boyd. This has now gone through seven editions. The term "degrees of freedom" does not appear in the index, but the concept occurs constantly throughout the book.

Goulden, C. H.: *Methods of Statistical Analysis.* John Wiley and Sons, Inc., 1939. See index.

Guilford, J. P.: *Psychometric Methods.* McGraw-Hill, 1936, p. 308.

Mills, F. C.: *Statistical Methods Applied to Economics and Business.* Henry Holt & Co., 2nd ed., 1938. See index.

Rider, P. R.: "A survey of the theory of small samples." *Annals of Mathematics.* Vol. XXXI, 1930, pp. 577–628. Published as a separate monograph by Princeton University Press, $1.00. Gives geometric approach to sampling distributions.

Rider, P. R.: *An Introduction to Modern Statistical Methods.* John Wiley and Sons, Inc., 1939. See index. While there is no general explanation of the meaning of degrees of freedom, this book gives a careful and detailed explanation of how the number of degrees of freedom is to be found in a large variety of situations.

Snedecor, G. W.: *Statistical Methods.* Collegiate Press, Inc., 1937, 1938. See index.

Snedecor, G. W.: *Calculation and Interpretation of Analysis of Variance and Covariance.* Collegiate Press, Inc., 1934, pp. 9–10.

Tippett, L. H. C.: *The Methods of Statistics.* Williams and Norgate, Ltd., 1931. One of the few attempts to treat the concept of degrees of freedom in general terms, but without geometric background, is made on pages 64–65.

Yule and Kendall: *Introduction to the Theory of Statistics.* Charles Griffin & Co. London, 1937, pp. 415–416, 436.

STUDIES IN THE PSYCHOLOGY OF MEMORIZING PIANO MUSIC: II. A COMPARISON OF MASSED AND DISTRIBUTED PRACTICE

GRACE RUBIN-RABSON

New York City

The step-by-step construction of a valid and useful technique of learning and memorizing keyboard music occasionally entails the experimental repetition of long-settled psychological problems. So thorough have been the investigations into the relative efficiencies of massed and distributed learning that McGeogh[6] has already summed up the results in the law that "some form of distribution is always superior to massed practices. It holds for virtually all experimental techniques and for all subjects and materials." This inclusive generalization derives from the conclusions of numerous studies in the areas of nonsense, motor, and logical learning material.

The classic experiments in memorizing nonsense material of Ebbinghaus[2] and Jost[5] demonstrated the superiority of distributed over massed learning. Perkins[7] emphasized this finding and concluded in addition that a reading every third day provided an optimum spacing for learning such material.

Edwards[3] stressed the value of spaced learning for small amounts of sense material. Austin[1] corroborated this result by using larger amounts of logical material. He summarized as follows: "No matter in what way the material was learned, the loss in the first two weeks was very rapid. It is for retention after longer intervals that the value of divided repetitions is particularly noticeable."

For achieving motor skills, as in learning typewriting, Pyle[8] favored distributed learning since "after a certain amount of practice, it seems that time must elapse before practice is beneficial." . . . "Distributed practice is the best, but the concentrated practice brings in good returns, and if one is in a hurry to acquire a skill, the diminished returns for concentrated practice need not be considered prohibitive." . . . "While inferences from one type of experiment to another are dangerous, the conclusion that the results here obtained would apply to learning to play the piano is probably warranted."

It is this hazard in making inferences from one learning area to apparently related areas that prompts the present experiment. The analogy between typewriting and playing the piano is speedily

exhausted once the space relationships on either keyboard have been determined.

The memorizing of piano music involves motor memory, logical memory, and, unfortunately, (from empirical observation) a memory for nonsense material as well.

The importance of emphasizing the logical aspects of the material in the first stages of the learning has already been conclusively demonstrated.[9] When the score has been exhaustively analyzed for intrinsic relationships of melody, harmony, and structural details preliminary to the keyboard practice, the amount of retention achieved by the analysis is markedly greater than that produced when the learning has been confined exclusively to the keyboard.

Because of the possibility that music students may make skillful analyses of new material with little awareness of the tonal effect, one procedure in the same experiment included study of the musical score while listening to several recorded repetitions of the composition. This variable effected no differences.

In a subsequent experiment,[10] and subordinate to the main issue, the problem of tonal awareness was again attacked, but in a more practical manner. The analytical study period was interrupted midway by several repetitions by the subject at the keyboard; the first half of the study period prepared him adequately for smooth performance; the second half allowed him to make further observations of the material with the tonal image in mind and the hand patterns already somewhat formed. The results were again negligible; but where small differences existed, they were in favor of the uninterrupted study period. It might have been assumed that the allotted number of intermediate repetitions would have reduced the total learning repetitions by a like number. But such was not the case. On the contrary, these intermediate repetitions seemed to exert only a destructive influence on the continuity of the analytical study.

The primary purpose of this experiment[10] was a comparison of the unilateral *versus* the coördinated approach—to determine whether any advantage derived when the coördinated trials were preceded by separate hand trials as contrasted with learning by coördinated trials alone. Had the experimental design been limited to the customary criteria of the number of required learning and relearning trials, no differences in the two approaches would have been evidenced. After the relearning, however, the subject transcribed the musical score from memory, and here real differences were apparent. The unilateral

approach insures a firmer grasp of the material, especially of the material in the left hand. Though equally fluent in performance after approximately the same number of relearning trials, the coördinated approach hides a basic weakness in the apperception of the details of the music.

It is likely that the experimental procedure obscured another possible superiority in the unilateral method. Each hand was brought to fluent memorized performance before the first coördinated trial; hence a certain amount of negative interference operated during the first coördinated trials. In spite of this interference, however, there were no reliable differences between the total number of learning trials required by either method. Had a coördinated trial been interpolated between each right-left sequence, this interference might have been eliminated and the total number of learning trials by the unilateral method been considerably reduced. The experimental technique in the present instance has been improved by this modification.

THE PROBLEM

In investigating procedures for the piano student, it is advisable, as far as is compatible with good experimental technique, to keep laboratory conditions similar to real practice situations. Certain conditions have become standardized in music study. The student practises every day and appears once or twice a week for the teacher's guidance. It seems impractical, therefore, to investigate the effect of a rest period of more than twenty-four hours duration; or the effect of spacing the repetitions in such small numbers that the finished smooth performance would require more than a week's preparation. The question is pertinent, then, in this form: does massed practice ending in smooth memorized performance yield better or worse retention after a lapse of two weeks than some form of distributed practice? Will the distributed practice be affected by confining it to one day as compared with distributing it over two days?

THE EXPERIMENT*

Three methods of practice were contrasted, of which one was massed practice and two were distributed as follows:

(*A*) Massed learning. Ten trials were allowed for the right hand, ten for the left hand, and ten coördinated trials. If at the end of the

* The experiment was conducted in the studio of the experimenter in New York City during the Summer of 1938.

allotted number the criterion of performance had not been achieved, the trials were continued.

(*B*) Distributed learning—two practice periods in one day. Five trials were allowed for the right hand, five for the left, and five coördinated trials. About one hour later, the learning was completed, using the second five trials, and continuing, if necessary, to reach the standard required.

(*C*) Distributed learning—two practice periods in two days about twenty-four hours apart. The distribution of trials was the same as for method *B*, with the altered time interval.

For all three methods, the standard of completed learning was one smooth memorized performance by the right hand alone, one by the left hand alone, and one smooth coördinated performance.

The total number of trials allowed for the learning was thirty—ten for the right hand, ten for the left hand, and ten coördinated trials. This number was estimated from the average number of trials required in previous experiments for somewhat easier material. How accurate this estimate proved to be can be seen by the very small number of extra learning trials that were needed beyond the given number. The ten trials were divided into two equal parts for the distributed learnings. Many of the learnings were, of course, accomplished in less than the allotted number. Those beyond the learning point were treated as over-learning trials; when necessary to reach the required standard, the trials were carried beyond the given number, and treated as "extra" trials.

Though not directly affecting the problem of massed and distributed practice, in order to build up a step-by-step procedure in memorizing piano music, the valid conclusions of previous experiments have been incorporated into the methodology of this one. Therefore, each experimental composition was studied and intensively analyzed before any work at the keyboard; the memorizing was a function of the learning; all keyboard trials were executed in a recurring sequence of right hand alone, left hand alone, and both hands together.

Because all the learnings followed the same general pattern, the directions to the subject were simple and general. Each composition was studied away from the keyboard for seven minutes. At the piano, no errors of note or rhythm were allowed. The subject was unaware whether he was to complete the learning at one sitting or whether he would be interrupted; hence his learning attitude was equally intense during the first repetitions, and the first five trials by any of the three

methods should show no differences. The keyboard procedure was always the same: One right hand trial, one left hand trial, one coördinated trial—this cycle was repeated throughout.

The subjects were warned against any rehearsal outside of the laboratory.

Relearning was accomplished exactly two weeks from the date of completion of the learning. At that time, there was no preliminary study and all the keyboard learning was effected at one sitting and by coördinated trials.

After the relearning, the subject transcribed the musical score from memory.

Experimental Materials.—Nine eight-measure periods were adapted from compositions covering a wide range of era and style. They were musically complete and of moderate though not equal difficulty, presenting some challenge to even the most capable subject. Table I

TABLE I.—VARIATIONS IN DIFFICULTY OF THE EXPERIMENTAL COMPOSITIONS AS SHOWN BY THE AVERAGE LEARNING AND RELEARNING TRIALS AND BY THE AVERAGE ERRORS IN TRANSCRIPTION

Composition No.	Learning trials	Relearning trials	Transcription score
1	20.8	8.3	11.8
2	25.0	9.0	25.0
3	19.4	6.2	11.8
4	21.4	7.9	10.8
5	13.5	6.0	10.2
6	20.7	7.9	14.8
7	18.4	7.1	8.3
8	25.3	10.7	6.8
9	20.2	8.0	12.1

shows the diversity among the compositions as evidenced by the average number of trials needed for learning and relearning and by the average errors in transcription.

Subjects.—The nine subjects, one man and eight women, were all experienced pianists. None had had less than seven years of formal piano instruction or less than two terms of training in theoretical music. From this minimum, training ranged to sixteen years of piano study and twenty-four terms of theoretical training. Table II shows some of the attributes of the experimental group. The subjects, with a mean age of twenty-two and four-tenths years were, according to Thorndike,[11] at the optimum learning age. A mean of eleven and two-tenths years

TABLE II.—MEANS AND RANGE OF THE AGE, MUSICAL TRAINING, AND EXPERIMENTAL ACHIEVEMENT OF THE GROUP OF SUBJECTS AS A WHOLE

	Means	Range
Age	22.4	19 –26
Piano study (years)	11.2	7 –16
Theoretical training (terms)	14.4	2 –24
Learning trials	20.5	13.4–24.3
Relearning trials	7.9	5.8–11.8
Transcription errors	12.3	1.6–28.3

of piano study and fourteen and four-tenth terms of theoretical training shows them to be persistent and serious music students. The range of thirteen and four-tenths to twenty-four and three-tenths learning trials, five and eight-tenths to eleven and eight-tenths relearning trials, and one and six-tenths to twenty-eight and three-tenths for the transcription score, indicates a wide range in individual ability.

Experimental Design.—Since musical ability comprises both measurable and immeasurable capacities, it is impossible to equate subjects on this basis. Nor is it necessary to equate them on this or any other basis if the experimental procedure conforms to the Latin-Greek square, in which all learning methods are rotated, and the other experimental variables of materials and order of learning are presented to the separate groups of subjects in ever-changing succession. With all the other variables operating equally, it is possible to analyze the influence of the methods themselves.

The nine subjects were divided into three groups of three each. Each group followed an individual Latin-Greek square which provided for the learning of three different compositions by three different methods. Therefore, each subject learned nine compositions; the group produced twenty-seven learnings, nine by each method. The three individual Latin-Greek squares became in turn parts of one large one and were themselves rotated. In this way, a total of eighty-one learnings, or twenty-seven by each method, was procured, with all the variables presented in every possible combination and order.*

* Due to the exigencies of time, several learnings were done during each laboratory session. In the *B* method, in which the learning was completed in two practice periods on the same day, one practice period was placed at the beginning of each session, the second at the end. There was, therefore, no rest period between the two practice periods, but an interpolation of other learnings. This was balanced by placing the second practice period of method *C* after the same number

ANALYSIS OF DATA

The Latin-Greek square lends itself readily to two basic statistical analyses. The economies of the methods may be evaluated by testing the learning and the relearning trials for the real or significant differences of their means, and for differences in accuracy in transcribing the material after relearning. Used in conjunction with the test of significant differences, an analysis of the total variance according to Fisher's[4] formula segregates the individual variables, both known and unknown, and assigns to each its relative contribution to the whole.

Learning Trials.—In all, thirty trials were allowed for the learning by each method, ten for the right hand, ten for the left hand, and ten coördinated trials. Table III shows the number of trials needed to reach perfection, (the number of trials past this point and up to ten may be considered over-learning), and the number of extra trials beyond ten needed to reach the learning standard. Consistently, the massed practice reaches perfection in the least number of trials, provides the most overlearning, but hovers between the distributed methods in the number of required extra trials. The *B* method (two practice periods in one day) consistently reaches perfection in fewer trials than the *C* method, (two practice periods on two days), provides more over-learning, but requires more extra coördinated trials. Apparently the learning is impaired between practice periods and requires several trials to restore it to its original state. After only one learning, it is, of course, more impaired after a lapse of twenty-four hours than after a lapse of only one hour.

Since for all three methods the first five trials were under identical conditions, an examination of the second five trials may prove fruitful. Table III shows more specifically the amount of impairment between practice periods, presenting the differences of the means of the methods, the Standard Error of the difference, and the ratio of the difference to the Standard Error.

In bringing the right hand to perfection, the second five trials of the *A* or massed-practice approach require significantly fewer trials than

---

of learnings on the following day. Should retroactive inhibition be operating in these two methods, then, it would operate to the advantage of method *A*, or the massed-practice method. In addition, method *B* would have some advantage over method *C* in that the second practice period of method *B* occurred at the end of the laboratory session, a favorable spot for retention. That none of these advantages accrued will be seen from the statistical analysis.

either of the distributed-practice methods. No differences exist in the number of extra trials needed. These economies appear again for the left hand; in addition, the *B* method is superior to the *C* method. No reliable differences inhere in the extra trials. For the coördinated

TABLE III.—MEANS OF TEN TRIALS REQUIRED TO REACH PERFECTION, AND SIGNIFICANCE OF DIFFERENCES OF THE MEANS OF THE SECOND FIVE TRIALS REQUIRED TO REACH PERFECTION

	Trials to perfection	SE difference	Ratio	Extras	SE difference	Ratio
Right Hand—Second Five Trials						
A-B	−1.33	.46	2.9	.04	.95	0.04
A-C	−1.45	.45	3.1	.34	.35	0.97
B-C	−0.22	.30	0.73	.38	.36	1.06
Means of Ten Trials (Right Hand)						
A	7.07	.40	....	.44		
B	8.33	.38	....	.41		
C	8.71	.26	....	.78		
Left Hand—Second Five Trials						
A-B	−0.96	.40	2.4	.10	.20	0.50
A-C	−1.92	.38	5.0	.41	.39	1.05
B-C	−0.96	.26	3.7	.51	.36	1.42
Means of Ten Trials (Left Hand)						
A	7.30	...	....	.29		
B	8.07	...	....	.19		
C	9.22	...	....	.70		
Coördinated—Second Five Trials						
A-B	−0.26	.31	0.84	.66	.91	0.72
A-C	−0.30	.29	1.02	.19	.80	0.23
B-C	−0.04	.18	0.22	.85	.85	1.00
Means of Ten Trials (Coördinated)						
A	9.30	...	....	2.93		
B	9.56	...	....	3.57		
C	9.63	...	....	2.74		

learnings, none of the methods yields any superiority. This more specific analysis emphasizes the conclusion already stated that several trials are needed to restore the learning to its original state after the interpolated time interval. That the coördinated trials show no differences may be due to the fact that one or two separate hand trials were sufficient to restore it to its original freshness.

When the total numbers of learning trials by the three methods are compared,* all apparent superiorities disappear. Table IV shows the means and standard deviations of the total learning trials. The *C* method, though requiring slightly more trials, produces the smallest fluctuation around the mean; the *B* method the largest fluctuation.

TABLE IV.—MEANS AND STANDARD DEVIATIONS OF THE LEARNING TRIALS AND THE SIGNIFICANCE OF THE DIFFERENCES OF THE MEANS

Method	Mean	SD	Difference of means	SE difference	Ratio
A, massed practice.....	19.8	4.6	*A-B* −0.7	1.3	0.54
B, two practice periods, one day..	20.6	5.2	*A-C* −1.4	1.1	1.3
C, two practice periods, two days..	21.2	3.4	*B-C* −0.7	1.2	0.58

TABLE V.—ANALYSIS OF THE VARIANCE OF THE LEARNING AND RELEARNING TRIALS AND OF THE TRANSCRIPTION ERRORS, SHOWING THE RATIO OF EACH VARIABLE TO THE RESIDUAL ERROR, FISHER'S ONE PER CENT VALUE, AND THE PERCENTAGE OF THE WHOLE DUE TO EACH VARIABLE

Variable	Fisher's one per cent value	Learning trials		Relearning trials		Transcription errors	
		Ratio	Per cent	Ratio	Per cent	Ratio	Per cent
Subjects.....	2.82	5.7	27.0	7.5	45	14.4	64
Methods.....	4.98	1.0	5.5	5.3	32	1.1	5
Compositions....	2.82	8.0	38.0	2.7	16	4.2	19
Order.....	4.98	5.1	24.0	.2	1	1.6	7
Residual error...	....	...	5.5	...	6	....	5

From Table IV it is evident that none of the mean differences are reliable. Judged by the number of trials required to reach perfection, no advantages accrue to any of the methods.

* In computing total learning trials, each right-left sequence was treated as the equivalent of one coördinated trial.

Table V presents an analysis of the variance which ascribes to each variable its proportion of the whole. Of the experimental variables, four are known: The differences among the subjects, the differences among the compositions, the effect of the order of the learnings on each other, and the differences among the methods. When these are accounted for, a residual error remains which is that part due to unknown, hence uncontrollable, factors. Of the total variance, the error accounts for only five and five-tenths per cent. The methods, already proved unspecific, account for the same amount. The considerable differences among the subjects (see Table II) and among the experimental compositions (see Table I) are responsible for twenty-seven per cent and thirty-eight per cent, respectively, of the whole. Because of increasing accommodation to the discipline of the experimental situation and because inherent in the methods themselves is the disposition of time of learning, the variable, order, bears a twenty-four per cent portion of the total variance.

Of the four controlled variables, only method does not show a ratio with the residual error comparable to Fisher's[4] one per cent value, a value which might be exceeded in random sampling from a homogeneous population once in a hundred trials.

Relearning Trials.—Table VI shows the means and standard deviations of the relearning trials of the three methods. The mean of

TABLE VI.—MEANS AND STANDARD DEVIATIONS OF THE RELEARNING TRIALS AND THE SIGNIFICANCE OF THE DIFFERENCES OF THE MEANS

Method	Mean	SD	Difference of means	SD difference	Ratio
A, massed practice.....	9.26	4.8	*A-B* −1.93	1.09	1.77
B, two practice periods, one day..	7.33	2.8	*A-C* −2.11	1.03	2.05
C, two practice periods, two days..	7.15	2.2	*B-C* −0.18	0.70	0.25

method *C* shows that practice on two separate days provides the best retention. In addition, of all the methods it shows least fluctuation around the mean. Second in amount of retention is method *B*, the other method of distributed practice. Least efficient is the method of massed practice.

This observation is confirmed by Table VI. The ratio of 2.05 indicates real differences between methods *A* and *C* in favor of the latter. Method *B* is reliably better than *A*, with a ratio of 1.77. The compositions learned by the *C* method were relearned fifteen days after

the first learning, whereas those by the *B* method were relearned only fourteen days after the first learning. The ratio of .25 between *B* and *C* proves that this extra day was ineffective.

TABLE VII.—GROSS RELEARNING TRIALS BY METHODS AND BY SUBJECT GROUPS

	Method *A*	Method *B*	Method *C*	Totals
Group I	61	60	63	184
Group II	58	52	51	161
Group III	131	86	79	296
Totals	250	198	193	

Though the method of massed practice is apparently least efficient, a closer analysis of the relearning trials yields an extremely interesting qualifying factor. Table VII presents the unaveraged relearning trials. It will be seen that Group I experienced little difference in the efficiency produced by the methods; Group II a difference of only seven trials between methods *A* and *C*. Group III is almost entirely responsible for the inefficiency of method *A*. From the totals in the right-hand column it is clear that Group II comprised the most capable subjects and Group III the least capable. When the nine subjects were ranked in order of ability as indicated by their average relearning trials, the subjects in Group III ranked seventh, eighth, and ninth. (The validity of this ranking is supported by a correlation of .72 between the relearning trials and the errors in the transcriptions.) Had this group learned and relearned the three most difficult compositions by the *A* method, it would be expected that the number of relearning trials would exceed the number of trials of the other methods which had utilized easier compositions. But the three compositions used in the *A* method by this group ranked fifth, second, and third, respectively, for difficulty. Furthermore, in the *C* method which required the least number of relearning trials by Group III, the compositions ranking eighth, sixth, and fourth, were used. (The reliability of the ranked difficulty of the compositions is strengthened by a correlation of .89 between the learning and the relearning trials.)

Group II, on the other hand, showing a plurality of seven by method *A*, used compositions ranking ninth, fourth, and eighth for this method, and the small difference is probably due to this fact.

Summarizing these observations, no differences are evidenced in method by Group I; for Group II, the slight difference seems to be due to the compositions themselves; but for Group III, comprising the

least capable subjects, a marked inefficiency inheres in the method, which is not due to the compositions; hence, this inefficiency must be a function of the capability of the subjects.

This conclusion is supported by the analysis of the variance of the relearning trials (Table V) where forty-five per cent of the whole is due to the differences among the subjects. The methods are now important, claiming thirty-two per cent of the total. Compositions, with sixteen per cent, are less significant than in the learning trials where the range was greater due to the inclusion of the separate hand trials in the total count. Unlike the original learning, all the relearning was accomplished by the massed procedure; therefore, order here is not a function of the method and is only one per cent of the whole. The six per cent for error is consistent in all the variance analyses.

The ratio for methods now exceeds Fisher's one per cent value while that for compositions barely reaches it. Order alone shows a negligible ratio.

Transcription Score.—Unlike other areas of learning, the fluency of memorized keyboard performance is not always dependable evidence of the precision and accuracy of the learning in deliberate recall. To test this precision, after the relearning each composition was transcribed from memory. These were scored for errors by a system in which omissions, additions, transpositions, infractions of rhythm, and the like were standardized. Scored once by the experimenter, they were checked again by an impartial assistant. Table VIII gives the

TABLE VIII.—MEANS AND STANDARD DEVIATIONS OF THE TRANSCRIPTION SCORES AND THE SIGNIFICANCE OF THE DIFFERENCES OF THE MEANS

Method	Mean	SD	Difference of means	SE difference	Ratio
A, massed practice.....	13.85	16.8	*A-B* −1.29	4.09	.31
B, two practice periods, one day..	12.56	12.4	*A-C* −3.11	3.60	.86
C, two practice periods, two days..	10.74	8.4	*B-C* −1.82	2.90	.63

means and standard deviations of the errors of the transcription score by the three methods. Method *C* insures the greatest accuracy, producing the fewest mean errors and the least fluctuation. But it is clear from Table VIII that these differences are insignificant. Though the retention value of the three methods differs, no appreciably greater accuracy is afforded by any one.

Most of the variance, (Table V) sixty-four per cent, is due to the varying capabilities of the subjects. The compositions are important

with nineteen per cent. The method, as already demonstrated, is not specific, and is only five per cent of the whole, as is the residual error. Some improvement in handling the writing problem was observed from day to day, as witnessed by the seven per cent for order.

Several explanations for the superiority of distributed practice have been advanced by other workers: The greater freshness after rest periods, the dropping out of errors, the length and difficulty of the material. The first two of these are eliminated in the present instance by the experimental procedure, the third has been ruled out in the statistical analysis.

The following hypothesis is tentatively suggested to explain the superiority of the distributed method for less able learners:

Inherent in musical structure is an organization and grouping of patterns usually independent of the demarcations of the measure; these patterns depend in part on the readiness with which certain keyboard units fall into hand groupings and in part on a grasp of intrinsic small patterns within the large patterns which may not be immediately discernible. Speed in learning and memorizing is directly related to the early perception of these groupings and patterns. For the less able learner, on first acquaintance with the whole configuration, these patterns may not emerge from their background, but must be "re-seen" in another situation after some familiarity with the material has been achieved. During the massed learning, no opportunity is present for this "re-seeing" so that the pattern is never clearly seen against its background and appears only in its first configuration. Such learning is, in a sense, dependent on repetition, and not on a logical reconstruction of the units of structure. It can, therefore, produce only a relative amount of retention.

SUMMARY AND CONCLUSIONS

Because of the hazard of applying inferences from one experimental area to apparently related areas, the present experiment was designed to retest the long-established superiority of distributed over massed practice. Though this superiority has been demonstrated in learning nonsense and logical material as well as motor skills, it can not be unquestioningly assumed that it will hold for memorizing piano music which, in some aspects, resembles all three.

For evaluating the advantages of bringing the learning to smooth memorized performance at one sitting as compared with distributing the practice on the same or the following day, the experimental situa-

tion has been kept as nearly as possible like a true learning situation. Few serious music students confine the daily practice to the study of one composition; therefore, the interpolation of other learnings between the distributed practice periods, though it introduces some element of retroactive inhibition, nevertheless more nearly approximates the daily practice routine. In the present instance, this variable was controlled and seemingly did not militate against the efficiency of the methods in which it operated.

Other workers have found certain optimum rest intervals in distributed learning. What those intervals may be for music learning is still unsettled; but because it conforms to practical usage, the shorter interval between the beginning and the end of one day's practice, and the longer interval between the practice periods of two successive days, have here been compared with a massed learning in the midst of other learnings.

The nine adult subjects were experienced piano students, ranging considerably in ability. Rotating all variables in every presentation, each subject repeated the experiment three times, using different experimental materials each time. The experimental design permitted an evaluation of the methods in terms of the materials, the capability of the subjects, and the length of the interpolated period.

The number of learning trials required by the three methods showed no differences, though the massed-trial study produced a greater degree of over-learning. The relearning trials demonstrated a real economy for the two distributed methods, the larger difference existing between the massed-trial approach and the two-day distribution. However, the two forms of distributed learning show no advantage in one or the other and the twenty-four-hour rest interval is unspecific in effect.

The number of errors made in transcribing the material from memory after relearning indicated no greater vividness or accuracy by any one of the three methods.

An analysis of all the variables in the relearning situation proves conclusively that, in this instance, the advantage of the distributed practice is in inverse relation to the ability of the subject.* It should, therefore, be recommended for less able learners. For the more able, either approach is equally productive.

Several explanations of the superiority of distributed practice have been advanced: One is that errors are permitted to drop out of the

* Learning ability has been found, in other experiments using some of the same subjects, to be almost independent of the amount of piano and theoretical training.

learning—in this case, no errors were allowed; another is the freshness and lack of fatigue for the new attack—but in these distributed learnings there was no real rest period due to the interpolated learnings; still a third is the nature of the material. The material used in this experiment was of considerable difficulty, some of it as intricate as many piano students are ever asked to memorize. In this case, however, the material was not a decisive factor, though extremely difficult music might be baffling enough to cause fatigue on an emotional basis.

A tentative hypothesis is suggested in explanation of the superiority of the distributed practice: The intrinsic patterns and groupings of the musical structure are readily perceived by the abler learners. For the less able, these patterns probably emerge from the general background only when they are "re-seen" a second time in a new situation. Since it is the apperception of these patterns which makes for real or logical learning, any procedure which re-presents the material in a new situation, regardless of the time interval between the presentations, will be more efficient than one that does not.

BIBLIOGRAPHY

1. Austin, S. D. M.: "A Study in Logical Memory." *Amer. J. Psych.*, 1921, Vol. XXXII, pp. 370–401.
2. Ebbinghaus, H. *Ueber das Gedächtnis.*
3. Edwards, A. S.: "Distribution of Time in Learning Small Amounts of Material." *Studies in Psychology*, Titchenor Commemorative Volume. Louis N. Wilson, Worcester, 1917, pp. 209–213.
4. Fisher, R. A.: *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh, 1934.
5. Jost, A.: "Assoziationsfestigkeit in ihrer Abhängigkeit von der Verteilung der Wiederholungen." *Zsch. f. Psych.*, 1897, Vol. XIV, pp. 436–472.
6. McGeogh, J.: "Experimental Studies of Memory." Robinson: *Readings in General Psychology.* Univ. Chicago Press, 1929.
7. Perkins, N. L.: "Value of Distributed Repetitions in Rote Learning." *Brit. J. Psych.*, 1914, Vol. VII, pp. 253–261.
8. Pyle, W. H.: "Concentrated versus Distributed Practice." *J. Educ. Psych.*, 1914, Vol. V, pp. 247–258.
9. Rubin-Rabson, G.: "Influence of Analytical Pre-Study in Memorizing Piano Music." *Archives of Psych.*, Col. Univ. No. 220, 1937.
10. Rubin-Rabson, G.: "Studies in the Psychology of Memorizing Piano Music: I. The Unilateral versus the Coördinated Approach." *J. Educ. Psych.*, 1939 Vol. XXX, pp. 321–345.
11. Thorndike, E. L., Bregman, E. O., Tilton, J. W., and Woodyard, E.: *Adult Learning.* Macmillan, New York, 1928.

THE RELATION OF RATIONAL TO GOAL INSIGHTS IN THE CURRICULUM

J. A. LYNCH

The Rice Institute

The problem which I am attempting to state in this paper is suggested by the relation of ideas to aims in Dewey's *instrumentalism* and by the interpretation which some of the writers of the Gestalt school of thought put upon goals in relation to rational insights. I am suggesting some possible criticisms of the assumption that aims or goals should always be emphasized in the planning of curriculum. I believe that there are at least some important exceptions to this rule which might have educational significance. I shall attempt to indicate briefly some of the possible effects of an over-emphasis upon goals.

To avoid confusion, I am giving special meanings to the terms rational insight and goal insight. I shall make no attempt to identify these concepts with the analogous concepts mentioned above. The point of view from which I am attacking the problem is similar in some respects to the Gestalt approach; but I am making some conscious departures from that position which I hope will not be interpreted as simply vague or erroneous restatements of it.

What I am calling a rational insight exists as a mature learning product in the form of a tendency to explain a given situation or class of situations in a typical way. It is the tendency to make use of a particular principle, formula, or a pattern of relationships whenever it is relevant to the problem at hand. It is acquired by practice in rationalizing a series of situations which need be alike only in so far as each one calls for the same type of explanation.

In opposition to a goal-insight, defined as the prospect of a particular goal-fulfillment, I am positing a learning product to be referred to hereafter as a value attitude. The latter, as a mature learning product, is expressed in behavior as the tendency to seek a certain goal or class of goals. Objection to this statement will be raised on the grounds that goal-seeking is conditioned largely by appetites and needs which are not learned. I have tried to answer that argument in another paper.[1] Appetites of various kinds are treated as parts of

[1] Lynch, J. A.: "A Criticism of Dewey's Theory of the Stimulus," accepted for publication by the *Philosophical Review*.

the stimulus rather than as parts of the response. I prefer to put into the stimulus everything which does not accrue from experience. This seems to me to be the best way to isolate the fundamental problems of education. I do not mean to imply by this statement that appetites, etc., should not be taken into account. They should be; but in the same sense in which other environmental factors are considered.

There are other types of learnings besides the two under consideration; but the limited problem does not implicate them.

If all the experiences which influence the formation of a particular personality could be assorted and classified according to their bearing on the learning process, there would be as many groupings as there are distinguishable personal tendencies. One of these hypothetical groupings can be expressed schematically as a summation, in this manner:

$$\sum_{i=1}^{i=n} a_i = P \cdot A_n.$$

This formula supplements the above definitions of the two learning products. If the term a_i represents any one of the practice situations which can be used to mature a rational principle of some kind, then A_n stands for the average expectancy in the way of accrued learning for the particular n practices of that series. The factor P is the personal factor. This can be treated as a constant, because the concern is primarily with a given environmental influence rather than with the problems of individual psychology. Social psychology can treat the personal factor as a constant. Conceiving the curriculum is a problem in social psychology in this sense, applying it involves individual psychology which must treat P as a variable. The latter problem is defined as method as distinguished from curriculum.

Also, A_∞ can be used to represent complete maturity for the relevant learning product. It can also represent the extent to which A generalizes, as,

$$A_\infty = \sum_{i=1}^{i=\infty} a_i$$

Maturity is defined as the stage at which this generalization to all possible members of the series takes place.[1] The concept of maturity

[1] This concept is essentially in line with ideas held by H. C. Morrison, C. H. Judd, Charles De Garmo, and Charles McMurry. I have treated the concept of

is fundamental to the argument. It admits, however, of some variation in interpretation.

If the formula above is reserved for series which lead to the clarification of a rational principle, the following variation of it can be used to represent the value-attitude type of learning product:

$$\sum_{j=1}^{j=m} b_j = P \cdot B_m.$$

In this series, b_j represents any situation which is defined and qualified by a particular goal or class of goal-fulfillments. B_m represents the value attitude at the stage of maturity corresponding to the given m practices. (The summation implies that practices can be unequal in their influence upon the learning product.)

Analogously, B_∞ can be used to represent the maturity of the corresponding value attitude. Maturity can also be defined as the stage at which further practice does not appreciably influence the learning product. It is convenient for the purpose of reducing the scope of the problem to apply the law of averages to b_j and a_i as well as to the personal factor. This makes it possible to get a perspective from the point of view solely of the number of practices in the series. This gives a better view of relative maturity on the part of the various learnings in so far as they are represented by practice series.

The foregoing does not throw any light upon the problem of relating these various matured or generalized learnings. The next step is to find some kind of a unity (1) between learnings of the same type and (2) between those of the two opposing types.

The fact that more than one learning practice can be exercised or practiced by the same practice series affords the basis of unity between learnings of the same type. That is,

$$\sum_{i=1}^{i=n} a_i = A_n' = A_n'' = A_n''', \text{ etc.}$$

For example, A' can represent the principle of oxidation, A'' the concept of chemical change, and A''' the concept of the uniformity of nature. If it is assumed that the more general a concept the more practices required to mature it, A_n' might be approaching maturity long before A_n'''; because A''' is the more general and, therefore, the

maturity more fully myself in an article entitled, "The Problem of Transfer," *The Journal of General Psychology*, Jan., 1939.

more subjective of the two.[1] But A''' is exercised by other series which bear no analogy whatever with a_i as interpreted in terms of A'. For instance, the law of the pendulum, the laws of motion, the atomic hypothesis, Charles' and Boyle's laws, as well as many other scientific insights, are matured and clarified by series which also suggest the idea of the uniformity of nature. Until this higher principle matures, however, these various rational learnings merely coexist; there is no unity in a strict sense, but when maturity of the more general insight is an accomplished fact, there is a synthesis in the sense that the particular learnings actually reënforce and support each other. In relation to each other, the law of the pendulum and Boyle's law might be viewed as examples of the manner in which natural law is expressed. In this sense, there is a very real sort of transfer[2] from one of these series to the other. There is also a very obvious sort of synthesis taking place at the stage of maturity for A'''.

Similarly, consistency between goals-sought results in hierarchies of value attitudes which mutually reënforce each other as the "higher purposes" become increasingly apparent in the general scheme. The argument supporting this can be anticipated from the analogy with the statement above concerning rational insights and their inner relationships.

This, however, leaves the personality divided into two main sections of learnings without any suggestion of how these might be related. The relationship between these two classes of learning products depends upon the way in which goals sought and rational insights are assumed to be related to each other in the learning situation. This is a crucial point in the argument.

This relationship can be made apparent by an examination of the first series considered,

$$\sum_{i=1}^{i=n} a_i = A_n$$

Here, for example, a_i consists of two distinct aspects: The first is the core meaning which defines the series; the second consists of the sensory elements, symbols, imagery, etc. which constitute the "parts" and which convey the meaning.[3] In a series of this type, defined by

[1] I believe a case can be made for this assumption.

[2] Lynch, J. A.: "The Problem of Transfer." *The Journal of General Psychology*, Feb., 1939.

[3] These "parts" have to be interpreted in their relationship to the core of meaning, whether that relationship is natural or artificial.

the rational insight which it suggests, I am assuming that ends can come and go as "parts" or elements. Therefore, a goal-situation is treated as an object, and the so-called "goal insight"[1] is conceived as the perception or intuition of this object. In the total situation, the goal may be more or less important as an element of content. It may, in some cases, be indispensable to the meaning. It should be treated as a possible element in the rational type of learning situation, but not as identical with it. That is, the goal is not the stimulus[2] in the sense in which the latter is defined.

On the other hand, in the series,

$$\sum_{j=1}^{j=m} b_j = B_m,$$

the axis is defined by the goals sought rather than by the idea expressed. In this case, ideas, rational insights, come and go throughout the series, while goals of a given class or family are constant; in this case, the ideas are radial, while the goal aspect is axial; whereas, in the previous example, the insight is axial while the goals are radial. But rational insights sometimes suggest goals in so far as the latter are needed to complete or clarify the situation in which the insight finds expression. Goals which cannot be attained out of relation to a cause-and-effect pattern of some kind are always couched in rational insights which lead up to their fulfillment. There are, however, highly subjective or spiritual goals which can be exercised through the use of archaic meanings having no connection with physical causation. This is apparent in one type of mystic experience.

The unity comprehending the value-attitudes and the rational insights, then, is apparent in the over-lapping of the two types of learning situations. It is suggested that this over-lapping is partial rather than complete. That is, the circle representing all rational learning situations would cover a large part but not all of the circle representing those of the opposite type.

If, as suggested above, repetition is treated as the only variable in the stimulus, the possible cross-sections of personality portray any number of combinations according to the degree of maturity reached by each particular learning. The major differences would, of course, result from the relative maturity reached by the more com-

[1] Or goal-intuition.

[2] See Lynch, J. A.: "A Quantitative Description of the Stimulus." *Psychometrika*, Vol. III, 1938, pp. 95–105.

prehensive learnings. But the greatest differences would seem to depend upon whether the volitional (value-attitudes) or the rational learnings are farthest advanced toward maturity.

For instance, if conscious effort is made to organize the curriculum around the type of learning situation represented by $\sum_{i=1}^{i=n} a_i$ instead of $\sum_{j=1}^{j=m} b_j$, it is conceivable that the maturity of rational learnings will outrun the maturity of the volitional learnings. In this case, the practical interest, represented by the dramatization of goals, is subordinated to the speculative insights. That is, goals are included only to clarify the ideas. On the other hand, if the emphasis is reversed and the rational insights are subordinated to the practical interests, the latter might be included in the plan of the curriculum only for the purpose of broadening the perspective around desired types of goals. In Dewey's terminology, ideas take on the status of means to an end; that is, they become tools or instruments. Instrumentalism aims at this result. The so-called integrated[1] program is of this variety; it means that the typical stimulus is of the $\sum_{j=1}^{j=m} b_j$ type, in which the axis is determined by goals sought rather than by rational insights. Rational idealism of any sort, whether the emphasis is subjective, objective, or transcendental, involves the assumption that the typical stimulus is of the $\sum_{i=1}^{i=n} a_i$ type. The latter leads away from the practical interests or it includes them incidentally as teaching aids.

Interest implies "tension toward a *goal*" only when a goal helps create the learning-situation. W. F. Book's[2] statement that "motivation is an increase of that tension" is a true statement only when the goal is involved as a part. The example which Book used of the students who changed their attitude toward educational psychology as soon as it was made a required course in Indiana University seems to prove this point. It is easier to get interested in educational

[1] *Integration* around an aim or problem as opposed to *concentration* around a principle (F. W. Parker) or around a subject (Tuiskon Ziller).

[2] R. H. Wheeler's *Readings*, p. 257.

psychology for its own sake than to intuit its relationship to the larger and more complex situation envisaged as a teaching career. The imposition of a goal upon a learning-situation sometimes changes the learning pattern altogether.

The relationship between the clarity of wholes and the clarity of parts[1] suggests that the desirable outcome is not of the either-or variety but one which demands both "strong purposes" and "clear ideas." This conclusion follows from the assumption implied above, which can be stated in this manner: If goal-situations are elements or parts in a total learning-situation, the clarification of the former attends the clarification of the latter.

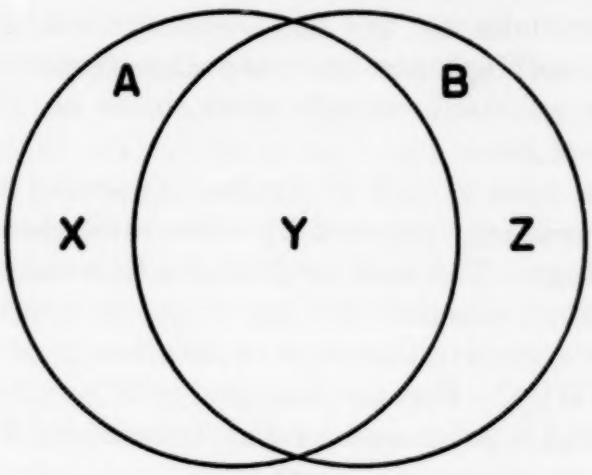

FIG. 1.

If all practice-situations for rational insights (all a_i's in the above formulas) contained goal-situations as "parts," and if all goal-situations existed only as "parts" in relation to such practice-situations, it would not seem to make much difference which emphasis is made; since the two kinds of practices would about equal each other, being the same in number.[2] But if this is not the case, it is possible for the extreme rationalist to build up practice-series, as curriculum content, some terms of which do not depend upon goals for their meaning.

---

[1] Wheeler's "Third Law" is suggestive.

[2] In so far as the time-interval between practices is a factor, it would likely favor the maturation of the type of learning product consciously selected. The time factor is obviously not as important from the Gestalt as from the association point of view. Time is a factor from the Gestalt point of view mainly from the standpoint of rebuilding depleted energies, appetites, etc., to their peak condition when these happen to function as parts in the learning situation.

In Figure I the sum-total of practice-situations for rational insights is represented by the circle with an area equal to *A*, and the sum-total of situations involving goal-situations by a circle whose area is *B*. The over-lapping part of the two circles is designated by *Y*, the part of *A* lying outside of *B* by *X*, and the part of *B* lying out side of *A* by *Z*.

The practice-situations out of which the practice-series are built up can obviously be selected: (1) From *X* exclusively; (2) from *X* and *Y* excluding *Z*; (3) from *Y* and *Z* excluding *X*; (4) from *X* and *Z* excluding *Y*; (5) from *Y* exclusively; (6) from *Z* exclusively; or, (7) from all three without discrimination.

(1) If the members of the practice series are selected entirely from *X*, the value attitudes are left undeveloped; because their practice-situations (goal-seeking) are systematically suppressed. The intellectual insights exercised, though clear, have no relevance to the practical affairs of life.

(2) Selection from *X* and *Y* can be interpreted as the condition for intellectual precocity followed by a less rapid development of the volitional learnings. The final result might be a balance between the two. A curriculum planned this way might be expected to produce individuals who remain rationalists or theorists even in the realm of the practical. Why? Because the number of practice-situations for the rational insights is favored by the exclusion of *Z* in favor of *X*.

(3) If the learning situations are consistently selected from *Y* and *Z*, the result expected is volitional precocity followed by a slower rational maturity, assuming that the educative process is continued long enough for the rational insights to mature finally. The last named condition might be expected to produce the type of individual who is intelligent in the pursuit of goals, but who, in a crisis, tends toward practical opportunism. This statement, like the previous ones, is based upon the assumption that maturity is the critical stage at which a learning product becomes effective in behavior.

(4) If the practices are chosen altogether from the two extremes, *X* and *Z*, the result is two simultaneous but unrelated educative processes, both of which exclude the practical. This might seem to be a rather improbable situation; yet, there are sensitive people whose dread of the practical might cause them to order their lives according to this fourth principle. For example, speculative science and mathematics might be balance against a highly subjective religious, artistic, or literary interest. There might be some kind of an emo-

tional union worked out between the two to meet the demands of a well-ordered life; or, the two might simply coexist.

(5) The practices taken from *Y* might be supposed to produce an approximately simultaneous development of both rational insights and value attitudes. This is based upon the assumption that it takes essentially the same number of practices to mature the rational insights as it takes for the value attitudes. This, of course, is not known to be the case; the assumption requires testing in some way.

(6) Selection from *Z* exclusively would develop highly subjective value attitudes out of experiences which are not related closely to real life. The goals sought, in this case, would have to be of a kind which are not easily connected with what is called reality. This is probably typified best by the mystic's desire for salvation of a kind which cannot be achieved through physical means. The aim involved is clear enough; it is dramatized, vicariously enjoyed, through the medium of poetical and philosophical expression; archaic meanings take the place of practice-situations of the more obvious variety.

(7) The last-named condition, based upon practices taken from all three regions, should be the ideal one for well-integrated development. Theoretically, the effect of this would be essentially the same as that resulting from the exclusive use of practices of the *Y* variety, except for the further possibility of integration on higher philosophical (rational) and religious (volitional or practical) levels. This statement rests upon the assumption that the range or variety of application is a factor in bringing about generalization, an assumption which I have attempted to justify in some other papers. On the other hand, there might be a disintegrative tendency in the central region if for any reason these extremes of the volitional and the rational fail to come together into a higher unity.

In conclusion, the question which I have tried to raise in the above discussion is really a very old one in philosophy. It concerns the freedom of the intellect. Rational thought can be subservient to this or that, religion, practical interests, or just our "interests." In a wide range of human experiences, that subordinate place is undoubtedly the proper one. But all aspects of human activity are entitled to a share of freedom. Rational thought is no exception. For example, the aesthetic interest in the fine arts does not thrive, does not reach its full development, in complete subserviency to advertising and other forms of commercialism. The excessive emphasis upon the practical seems to me to dwarf the rational side of education

in an analogous way. There is no substitute for the direct appeal of geometry, algebra, physics, chemistry, biology, history, biography, literature, etc., whether their practical value is clearly apparent or not. The mind of the child or youth thrives best on these materials as free. The above argument does not imply that education should necessarily be exclusively this sort of thing. The practical emphasis is entitled to a place in the curriculum commensurate to its importance. The point is that a one-sided theory of curriculum is likely to exclude the rational interest from its proper place. Value attitudes, conceived as learning products resulting from practice in goal-seeking, develop when the practical interests are free in this sense of freedom; that is, when they are emphasized. These value attitudes are necessary to successful participation in the practical affairs of life.

Dewey takes account of the speculative aspect of education by saying that the individual can become intrinsically interested in the things of the intellect as instruments,[1] like one might become a maker and a critic of tools which are viewed in relation to the uses to which they are put. This, in my opinion, excludes, or at least dwarfs, the speculative side, because the tool character of ideas is rarely obvious; besides, they are often discovered first and applied afterwards in situations which did not exist prior to their discovery.

[1] Particularly, Ch. XIII, *Essays in Experimental Logic*, Dewey attempts to answer the charges of over-emphasis on the practical.

THE INTER-RELATIONS OF VERBAL REASONING, NON-VERBAL REASONING, AND INTELLIGENCE

LEON BRODY

New York University

Tests of so-called verbal and non-verbal reasoning were constructed following the usual procedures involved in the standardization of mental tests. The verbal portion was characterized by the prevalence of words not only in the directions, but also throughout the reasoning items. The non-verbal part used language only in the directions, the reasoning items themselves being in picture or symbol form. Within each part, the reasoning was of two kinds—concrete and abstract, and in two forms—analogies and classifications. It was found possible, because of their nature, to have the non-verbal concrete items parallel completely in content the verbal concrete items. The non-verbal abstract and verbal abstract items are only arbitrarily assumed to be parallel, because of the common absence of real, concrete objects or things from the reasoning items they present.*

The tests went through three forms. The final form required an average of sixteen minutes for the actual testing time for the non-verbal part, and the same for the verbal part. The complete test was administered to fifteen hundred fourteen pupils in Grades IV through XII in four New York City schools. On the basis of this administration, the reliability of the subtests and of the entire test was determined and found to be entirely satisfactory.

VERBAL AND NON-VERBAL REASONING IN RELATION TO INTELLIGENCE

In referring to tests of reasoning, one inevitably is inclined to think of tests of intelligence. The tests of reasoning used here differ from the ordinary tests of intelligence in that only a limited number of functions or abilities is sampled. Furthermore, the tests have been so constructed as to facilitate an analysis of these specific functions or abilities, rather than to present an ordinary intelligence testing situation. However, there is always interest in a determination of the relationship between so-called reasoning tests and so-called intelligence tests. With this in mind, standard intelligence test scores were obtained for thirteen hundred one out of the fifteen hundred fourteen pupils tested,

* For a full description of these tests, see the previous article: "Comparable Tests of Verbal and Non-Verbal Reasoning: Their Construction and Application to Developmental Problems," *Jour. Ed. Psy.*, March, 1940, pp. 180–194.

and, in this way, correlations with intelligence were found for each of the four major types of reasoning tested, as well as for the entire test. (Four different intelligence tests were used for different levels. The Otis S-A, intermediate was used in Grades IV–VI; National, in Grades VII–VIII, Terman Group in Grade IX, and Otis S-A, Higher, in Grades X–XII.) These figures are as follows:

N-1301	*NVC*	*NVA*	*VC*	*VA*	Total battery
IQ	.40	.45	.52	.54	.53

Key: *VC*—verbal concrete reasoning. *NVC*—non-verbal concrete reasoning.
VA—verbal abstract reasoning. *NVA*—non-verbal abstract reasoning.

It will be noticed that the figures indicate substantial relationship in all cases, although the correlations are far from high; and that each of these correlations is definitely reliable when considered from the point of view of its probable error. In general, the verbal tests correlate more highly with intelligence as measured by the tests administered than do the non-verbal tests. Also *NVC* correlates about as highly with intelligence as *NVA*, and *VC* about as highly as *VA*. Finally *VA* correlates as highly with intelligence as does the entire test, which may be taken to mean that, as an indicative of intelligence measured by an ordinary intelligence test, *VA* is just as good as the whole test. Probably, the higher correlation of the verbal tests with IQ is due to the verbal nature of the intelligence tests used; namely, the Otis Self-Administering Test, the Terman Group Test, and The National Intelligence Test. It should be noted that a multiple correlation between both verbal tests and intelligence produces a coefficient of .56. In the case of non-verbal tests the coefficient becomes .46.

In studying the relation between scores on these tests and IQ, it must be remembered that IQ is a function of age. Therefore, to determine the true relation between intelligence or brightness and performance on these tests, it is necessary to hold constant the factor of age. This was done, and the following are the findings, the original correlation being indicated in parentheses:

IQ × *NVC*.age = .47 (.40)
IQ × *NVA*.age = .52 (.45)
IQ × *VC*.age = .65 (.52)
IQ × *VA*.age = .72 (.54)
IQ × total age = .70 (.53)
IQ × age = −.0226

The correlation between IQ and the combined *NV* tests was determined by multiple correlation. This turned out to be .46. Likewise, the correlation between IQ and the combined *V* tests was found to be .56. Partialling out the age factor, IQ × *NV* becomes .55 and IQ × *V* becomes .75.

It will be noticed that in all cases the correlation is raised. The increase is most in the case of the verbal tests, with *VA* and *VC* combined (*V*) being particularly outstanding. The latter, under this condition, correlates with IQ at .75, which indicates a significantly high relationship and the possibility of independent use as an intelligence test. The case of *VA* is almost as significant. In a way, verbal abstract reasoning represents an advanced form of man's mental processes and, therefore, this high correlation is to be expected. That the partial correlations are all higher than the zero order *r*'s is due to the fact that the uncontrolled factor of age, through its common relation to IQ and these tests, tends to lower the net correlation.

For the most part, the correlations shown here are not high enough to justify their use in prediction of the intelligence level of a particular pupil. However, in the case of the non-verbal tests, this does not mean that they do not measure another aspect or another kind of intelligence. Non-verbal tests of so-called intelligence rarely correlate highly with verbal tests of so-called intelligence. The Myers test seldom yields a coefficient of correlation of more than .55 to .60 with other tests. Many investigators take this to mean that these non-verbal tests do not really test intelligence; but this is not necessarily the case. It may be that tests which do not involve the interpretation of printed language are getting at another aspect or phase of intelligence. Furthermore, they are of definite practical value in many cases where the language factor has to be minimized or controlled. The writer feels that the present tests can be considered in this category. The small amount of time involved in their administration and their comparability with the verbal tests are peculiar advantages attached to their use. Lengthening and varying the tests may enhance their value.

Another approach to the problem of relations with intelligence may here be considered. In Table I a series of partial correlations is presented in which IQ is a factor. The partial correlation between IQ and *NVC* with *VC* held constant, reduces the original correlation from .40 to .01; while the partial correlation between IQ and *VC* with *NVC* being held constant reduces the original correlation of .52 to .37. In

connection with the major purpose of this section, these findings may be taken to mean that verbalization (or the verbal nature) of non-verbal test items of a concrete kind, contributes more to the relation between IQ and *NVC* than the non-verbal or pictorial factors ("picturing") in *VC* contribute to the relation between IQ and *VC*. Again, the verbality of the intelligence test used accounts in part for this fact.

TABLE I.—PARTIAL CORRELATIONS OF IQ AND VARIOUS SUBTESTS

IQ × *NVC.VC* = .01	IQ × *VC.NVC* = .37
IQ × *NVC.VA* = .05	IQ × *VC.NVA* = .31
IQ × *NVC.NVA* = .11	IQ × *VC.VA* = .16
IQ × *NVA.NVC* = .25	IQ × *VA.NVA* = .36
IQ × *NVA.VA* = .10	IQ × *VA.NVC* = .39
IQ × *NVA.VC* = .12	IQ × *VA.VC* = .22

Also, the partial correlation between IQ and *NVA* with *VA* held constant reduces the original correlation of .45 to .10; whereas the partial correlation of IQ and *VA* with *NVA* held constant reduces the original correlation of .54 to .36. This would seem to indicate that verbalization contributes more to intelligence as measured by the standard tests used than spacial and symbolic factors as such, *NVA* being verbalized to some extent. In this connection, it should be pointed out that *VA* contains verbal items which fit in with *NVA*, as the following two classification items indicate:

1. lofty	1. segment
2. above	2. portion
3. below	3. piece
4. high	4. time
5. up	5. part

That *VA* is related to *VC* and *VC* to *VA* in their relationship to IQ is indicated by the partial correlations with IQ in which each is in turn held constant. The new correlations obtained will be seen to be considerably lower than the original ones. This means a common ground for the two functions, whether it be due to similarity in test structure or to a verbal factor or to both. It is perhaps less correct to distinguish between structure and content than to speak of the set-up or testing situation as a whole.

That there is a common factor of verbality as such, apart from the nature of the content, (concrete or abstract), is suggested by the fact that the correlation between IQ and *NVC* is reduced to .01 when *VC* is held constant, and to an equally low level; namely, .05, when *VA* is held constant. We have the same situation when *VA* is held constant, and then *VC*, in ascertaining the relationship between IQ and *NVA*.

In connection with the correlation between IQ and *NVA*, if *VC* is held constant, we get a partial correlation of .12, whereas if *NVC* is held constant, we get a partial correlation of .25. This would seem to indicate that verbality or verbalization contributes more to the relationship between IQ and non-verbal reasoning of a so-called abstract nature, than does non-verbal reasoning of a concrete nature. This seems to be the case also, although not as significantly, with the partial correlation between IQ and *NVC* when *VA* and then *NVA* are held constant. The reverse, however, does not hold true; namely, that non-verbality or "picturing" is very essential to the relationship between verbal reasoning and intelligence. Other partial correlations point to the fact that verbal factors have much more to do with the relationship between intelligence and verbal reasoning than do non-verbal factors.

One way of corroborating some of the findings presented above is through use of multiple correlations (Table II). If one studies the combined effect of *NVC* and *VC* in relation to IQ, one will find a correlation of .53. Since the correlation between IQ and *VC* is .52, a possible conclusion is that *NVC* adds nothing to the relation between *VC* and IQ. This can be accounted for. *NVC* and *VC* are parallel tests and, as we have seen, *NVC* is probably verbalized to a definite extent. Therefore, we would expect little change in the original correlation. This, of course, is theoretical, since numerous factors may enter into a multiple correlation. Similar conclusions in this section are subject to the same limitation.

TABLE II.—MULTIPLE CORRELATIONS OF IQ AND VARIOUS COMBINATIONS OF SUBTESTS

R-IQ (*NVC* × *VC*)	− .53	*R*-IQ (*NVA* × *VA*)	= .55
R-IQ (*NVC* × *NVA*)	− .46	*R*-IQ (*NVA* × *VC*)	= .53
R-IQ (*NVC* × *VA*)	− .54	*R*-IQ (*VC* × *VA*)	= .56

The combined contribution of *NVC* and *NVA* to IQ produces a correlation of .46, which again indicates no significant addition to the original correlations between IQ, measured by the tests used, and each of these variables. In other words, they may have a factor in common which particularly seems to contribute to intelligence. This common factor, under the reasoning followed previously, could be a verbal factor, arising from verbalization. Other multiple correlations show no significant changes or departure from the original correlations.

[All the correlations mentioned in this section are probably higher than pictured because these correlations involve not one intelligence test, but three, which means a greater variability and increased

difficulty in obtaining high correlations. Also it should be pointed out here that no correlation has been corrected for attenuation. Significant increases might thus be obtained.]

VERBAL AND NON-VERBAL REASONING IN RELATION TO EACH OTHER

For the purpose of this paper, verbal reasoning has been taken to refer simply to performance on those tests which involve materially the use of words, and non-verbal reasoning has been taken to include performance on tests that are characterized by the absence of words from their basic content. It must also be pointed out that words and sentences are used in the test directions of all the tests; and that it has not been taken for granted that pictorial or symbolic subtests are only perceptual or that verbal subtests are only linguistic. The designations used are not defended. It is the purpose of this paper to be concerned only with interrelations of types of reasoning, whatever be the names assignable to any of these functions.

Similar considerations should be had with respect to the types of reasoning known as abstract and concrete. For present purposes, concrete reasoning has to do with thought problems revolving around real, tangible objects or things. Abstract reasoning is in this paper used in a negative sense; namely, to refer to reasoning of a type that does not deal with actual, tangible or concrete things. *NVA* is predominantly symbolic and involves relationships that are frequently, although not exclusively, spatial, presenting situations which have previously been unencountered as a whole, and even in part. *VA* presents more or less intangible life situations in verbal form, some of which are spatial. At least parts of these situations have been previously encountered.

Table III presents the inter-correlations of the major variables for all cases. It is to be noted, first of all, that the correlations are all fairly high and consistent, regardless of which pairs of subtests are used in correlation purposes. In the second place, the probable error of each correlation is very small, guaranteeing the reliability thereof. Thirdly, each subtest correlates very highly with the entire test, *VC* being outstanding in this respect. In other words, *VC* is a more typical test than any of the others insofar as it measures an ability which is common to all four sub-tests. It is the best measure, or the most valid measure, of the function with which the entire test is concerned.

Among the individual subtests, the greatest intercorrelation is between *VC* and *VA*, *r* being equal to .84. Table IV, which furnishes a

TABLE III.—INTER-CORRELATIONS OF MAJOR VARIABLES FOR ALL CASES

N-1514	*NVC*	*NVA*	*VC*	*VA*	Total
..	1	2	3	4	5
1	..	.73	.76	.69	.88
2	..	...	.73	.72	.88
3	..	...	...	.84	.93
4	..	...	...	...	.91

subjective factoral analysis of the reasoning tests by pairs, indicates that these two tests are particularly distinguished by the identity of their content structure. This correlation is reliably higher than the correlation of .76 between *NVC* and *VC*, which, according to the chart, are distinguished by the identity of their content nature. Off-hand, this may be taken to be of significance in serving to indicate that content structure (verbality, in this instance) determines correspondence between two tests more than content nature. This conclusion on the relative contributions of form and content may be limited by several considerations. *NVC* and *VC* were given on different days, whereas *VC* and *VA* were given on the same day, thus involving the question of the mental set of the pupils. Also *NVC* and *VC* each involved the same amount of time, whereas *VC* and *VA* differed slightly in time. Finally, *VC* and *VA* are identical in test structure, whereas *NVC* and *VC* are only similar in test structure.

TABLE IV.—SUBJECTIVE FACTORAL ANALYSIS OF REASONING TESTS BY PAIRS

	Test structure (mechanical set-up)	Content structure (pictorial, verbal, or symbolic)	Content nature (similarity of reasoning items)	Time (duration of test)	Sitting (same or different hour)
NVC-VC	*s*	*d*	*i*	*i*	*d*
NVA-VA	*s*	*d*	*d*	*i*	*d*
NVC-NVA	*i*	*i*	*d*	*d*	*i*
VC-VA	*i*	*i*	*d*	*d*	*i*
NVA-VC	*s*	*d*	*d*	*d*	*d*
NVC-VA	*s*	*d*	*d*	*d*	*d*

i—identical or almost identical.
s—similar but not identical.
d—different.
?—similar only in that both are non-verbal, objectively speaking.

If the conclusion is justifiable in spite of the considerations mentioned, it is in direct contradiction to the finding of Davey,[1] to the effect that the group factor in her semi-parallel verbal and pictorial tests was introduced by the content of the tests, rather than by their form. It might be said, however, that form distinguished from content is a purely theoretical abstraction.

At any rate, it seems that both content structure and content nature (or both form and content) are responsible for the high inter-correlations. Usually, such high inter-correlations give evidence in opposition to the point of view that has sometimes pictured verbal and non-verbal abilities as independent. Stephenson,[3] using non-verbal and verbal subtests that did not "parallel" one another, found a high correlation amounting to .82 for a summed correlation for many subtests of both kinds. This is in agreement with the present finding, for, by means of multiple correlation, *NV* and *V* are found to correlate at .80.

The present problem of inter-relationship lends itself to two approaches by means of partial correlations. One is through the determination of significant partial correlations among the four major variables, and the second, an indirect way, is to use a fifth variable as a control, holding constant one of the four variables when correlating another one of them with the fifth variable. The latter approach will be considered first. (See Table I.)

The fifth variable mentioned refers to the intelligence quotients determined by standard intelligence tests. The correlation between IQ and *NVC* is .40. Now, if one correlates these two factors while holding *VC* constant, r becomes .01, which is significant for its small size and great difference from the original correlation. An interpretation of this is that *NVC*, during the course of the test, is verbalized to a definite extent. This may indicate that there is a group factor or a general factor of a verbal nature among these verbal and non-verbal tests. In support of such a possible conclusion is the fact that the correlation of .52 between IQ and *VC* is reduced only to .37 when *NVC* is held constant. In other words, while there may be a specific *NV* factor, this evidence favors the existence of a more important group or general factor of a verbal nature.

Again, the correlation between IQ and *NVA* is .45, and if one partials out *VA*, it drops to .10, indicating that *NVA* may also be verbalized. On the other hand, the correlation of .54 between IQ and *VA* is decreased only to .36 when *NVA* is held constant, perhaps

further justifying the previous statement concerning the nature of a common factor. If, instead of holding *VA* constant in determining the relationship between IQ and *NVA*, *VC* is held constant, r is found to be equal to .12. This gives very much the same implications as the previous partial correlation of .10.

Another way of arriving at the same conclusion is through determining the combined contribution of *NVC* and *VC* to IQ. This multiple correlation turns out to be .53. Since the correlation between IQ and *VC* is equal to .52, this may be taken to mean that *NVC* adds nothing to the relationship. The reason for this may well be that *NVC* is verbalized, making *NVC* and *VC* very similar so far as actual reasoning processes are concerned. Likewise, the combined contribution of *NVA* and *VA* raises the correlation with IQ only to .55, the correlation between IQ and *VA* being almost exactly the same; namely, .54.

The second approach to this problem is through partial correlations involving the four major variables. For example, there is a correlation of .69 between *NVC* and *VA*. If *VC* is partialled out, this correlation drops to as low as .14, indicating that some kind of verbality may account a great deal for the "commonness" in ground of *NVC* and *VA*. Of course, such interpretations cannot be carried too far because of the complexity of the relationships existing.

TABLE V.—PARTIAL CORRELATIONS BETWEEN THE VARIOUS SUBTESTS

$NVA \times NVC.VC = -.40$	$NVC \times VA.NVA = .33$
$NVA \times NVC.VA = -.46$	$NVC \times VC.NVA = .48$
$VC \times VA.NVA = -.65$	$NVC \times VC.VA = .46$
$VC \times VA.NVC = -.68$	$NVA \times VC.VA = .35$
$NVC \times VA.VC = -.14$	$NVA \times VC.NVC = .40$

In the case of most other partial correlations the effect mentioned is not as great, although still sufficient to be significant. This is due to the fact that the only "true" parallel tests are *NVC* and *VC*. However, partialing out *VA* from the correlation of .73 between *NVA* and *VC* reduces the latter to .35. Holding *VC* constant in the case of the correlation between the two non-verbal tests reduces the latter from .73 to .40, and doing the same with *VA* reduces the correlation to .46. Compare with this the fact that holding *NVA* constant in the case of the two verbal tests reduces the correlation of the latter from .84 to .65; and doing likewise with *NVC* reduces the correlation to .68. The former reductions obviously are considerably greater

than the latter. All this, perhaps, points to the dominance or at least presence of a common factor marked by verbality.

SUMMARY AND CONCLUSIONS

(1) As measured by these tests, the four types of reasoning (non-verbal concrete, non-verbal abstract, verbal concrete, verbal abstract) do not individually seem to involve the same things measured by ordinary, verbal intelligence tests. Verbal concrete and verbal abstract combined approach the latter to a significant extent, probably because of common verbality. Non-verbal concrete and non-verbal abstract do not. This does not, however, mean that the verbal tests are better measures of intelligence as such, than the non-verbal tests. It may be that the non-verbal tests are getting at another aspect of intelligence.

(2) As determined by the tests used, the intercorrelations of the four functions measured are all fairly high and consistent, ranging from .69 to .84. By multiple correlation, the verbal and non-verbal tests, as such, correlate at .80. All this implies that the verbal and non-verbal abilities tested may not be independent, although the contrary is frequently claimed. Many factors account for the relationships existing between the specific functions, among them being test form and content. It is possible that a major common factor causing the high intercorrelations is verbality of a kind. Statistical approaches by means of multiple and partial correlations, aided by introspection, may be taken to indicate a tendency to verbalize so-called non-verbal items.

REFERENCES

1. Davey, Constance M.: "A comparison of group verbal and pictorial tests of intelligence." *British Journal of Psychology*, General Section, Vol. XVII, 1926, pp. 27–48.
2. Smith, G. M., Jr.: "Group factors in mental tests similar in material or in structure." *Archives of Psychology*, No. 156, 1933. New York: Columbia University, 1933.
3. Stephenson, William: "Tetrad differences for verbal subtests relative to non-verbal subtests." *Journal of Educational Psychology*, Vol. XXII, 1931, pp. 334–350.

IS ABILITY RANDOM OR ORGANIZED?

C. SPEARMAN

There has just been published a remarkable book by Professor G. H. Thomson, entitled the *Factorial Analysis of Human Ability*. Assuredly, it is in many respects an extremely fine work. Nevertheless he himself will be the first to agree with me, that everything on the subject has not been said even yet. In particular, I still feel constrained to touch on certain matters not quite free from contentiousness.

Frankly speaking, there has existed between him and me a very long opposition; one might almost say, a feud. However both of us —I hope and trust—have all the time been steadily progressing. And certainly this double movement has been along convergent lines. So far has such reconciliation gone, indeed, that by this time I find accepted in his expositions almost everything which I have ever regarded as my "main" theory; this comprises wholly and solely the verifiable facts of observation or of mathematics. This main theory I have always sharply distinguished from what I have called the "sub-theories," dealing only with hypothetical and not directly verifiable interpretations. The said main theory includes among other things, I may add, much the greater part of my *Abilities of Man*.[1]

Passing over to the sub-theories in these, there does seem to remain some difference between the two of us that still needs elucidation. Even here, however, we seem to be in harmony up to a certain point. For I myself, no less than Thomson, accept the hypothesis that the observed test-scores, and therefore their correlations, derive originally from a great number of small causes; as genes, neurones, etc.[2] Indeed this much seems to be accepted universally. We only disagree as to the way in which this derivation is to be explained. For my part, I have for this purpose had to invoke the fact that (under statable conditions) the correlations between any set of abilities become "hierarchical," in which case every test-score can be divided into two statistical factors which are, respectively, general and specific. I have

---

[1] To be precise, it includes the whole of two hundred sixty-six pages, whereas the sub-theories are allowed only forty-nine. The clarifying suggestion has been made by Thouless, that the general factor as statistically determined in the main theory might be denoted by "*g*," whilst this same factor as psychologically interpreted in any sub-theory could be distinguished as "*G*." See *The Study of Society*, ed. by Bartlett, Ginsberg, Lingren, 1939, art. by Blackburn, p. 163.

[2] See my article in this Journal, Vol. XXI, 1931, p. 158.

sought to interpret this statistical fact in hypothetical psycho-physiology (invented more sepecially to account for the phenomenon of mental span). I have suggested that the said multitudinous small causes jointly contribute to some or other general "power," "energy," "potential," or, above all, Lashley's "mass action." But Thomson for his explanation, although arriving at the same statistical general factor ("*g*"), has found no need for introducing any hypothetical energy. Instead, he has taken the derivation of "*g*" from the said multitudinous small causes to be sufficiently explained by supposing these to have been submitted to the operation of "random sampling,"[1] or "leaving the rest to chance."[2]

But herewith the fat is in the fire. Does or does not this operation of "random sampling" suffice to account for all that is definitely known? Or is it at least compatible therewith?

Now, up to a certain point, and this a fundamental one, the answer must indeed be in the affirmative. Random samples drawn from a very large number of elements—Thomson suggests the existence of something like a hundred thousand millions[3]—will really and truly produce a statistical "*g*" satisfying the principal criterion at issue (that of "hierarchy" or the "zero tetrad-differences"). Indeed, the satisfaction will not be merely as Thomson indicates, approximate; it will be as good as perfect. Moreover this fact, far from ever being denied by me, was given by Maxwell Garnett and myself as early as 1914 in what seems to have been the earliest suggestion of sampling at all.[4]

So far as concerns the production of hierarchy, then, the sampling theory has an easy game. But unfortunately it also has other products that are less satisfactory. Outstanding is the well-known fact that all very large samples must in general tend to have one and the same set of mean values (to wit, those which exist in the universe whence the sample is drawn). In such case, every individual should in every test obtain one and the same score; which is, of course, contrary to

[1] He reiterates and italicizes that he is speaking of "a large *random* sample." (*Brit. J. Psych.*, 1935, p. 72.)

[2] For instance he writes in 1919 that if "the mind has two (or even one) levels at which it can operate then, *leaving the rest to chance*, this theory will give hierarchical arrangement of correlation coefficients in the overwhelming majority of cases." (*Brit. J. Psych.*, IX, p. 341.)

[3] See his book p. 51.

[4] Spearman: *Psych. Rev.*, 1914. Garnett, Maxwell: *Brit. J. Psych.*, 1920. Spearman: "The Sub-structure of the Mind," *Brit. J. Psych.*, 1928.

observation. Equally inacceptable is the analogous corollary of the sampling; namely, that every test should have approximately the same constitution.

Still more forcible than these statistical objections to the assumption of random sampling are its psychological troubles. Surely, anyone who has spent hours and days refining a single mental test can never be happy with the doctrine that its constituents are merely put together at haphazard. And most cogent of all remains the objection raised by Slater:

> If *g* is an average or total of a man's bonds, and a *g* test a random sampling of them, the best test of *g* must be the one . . . which uses as much diverse material and presents it in as many diverse ways as possible. . . . Evidence suggests that this is not so. . . . In practice the development of *g* tests has been accompanied by a progressive narrowing of the types of problems, materials and methods of presentation used.

Such being in substance the chief objections raised time after time to the sampling theory, what has been the reply? This too has been always the same. It has been to the effect that the word "sample" is being *misunderstood*. Thus when Garnett and myself first urged these objections, Thomson attributed them to our having "adopted an entirely different meaning of the word 'sampling.'"[1]

In 1932 a great advance was attempted by Brown and Stephenson. These authors, besides submitting the objections to exact experimental verification, hoped to escape further wrong interpretations of the term "sample" by following that which had just been adopted in a new and admirable research of Blackie. The prospect seemed fair enough, for Blackie not only had once been a student of Thomson, but also in this very research had enjoyed his collaboration. When, however, the results obtained by Brown and Stephenson proved to be again adverse to the sampling theory, Thomson here too accounted for the failure on the ground that "the adoption of these equations (of Blackie) means giving up entirely what I had meant by sampling."[2]

Thereupon Brown abandoned the whole problem as being hopelessly:

> handicapped by the inadequacy of the formulae hitherto brought forward in connection with the Sampling Theory.[3]

---

[1] *Brit. J. Psych.*, 1935, p. 66.
[2] *Brit. J. Psych.*, 1935, p. 70.
[3] *Brit. J. Psych.*, 1935, p. 319.

Nevertheless, subsequent investigators have not ceased to urge these same old objections—only to meet with the same inconclusive response.

After such long and unprofitable cross-purposes, we turn in pleasant anticipation to Thomson's new volume. And here in good truth we do at last find some notable comfort.

At first sight, indeed, he might seem to leave all the old principal discords still unresolved. Thus, he continues to maintain that the factor-scores of each individual in each test constitute a "sample" out of all possible factor-scores. Similarly, he still seems to hold that the factor-loadings in each test are a "sample" of all possible factor-loadings.

On closer inspection, however, his new book presents a very remarkable innovation. He no longer contents himself with a bare disavowal of the meaning given to a "sample" by his opponents. He now goes on to supply his own definite interpretation of the term. He writes: a sample means "some but not all" of the elementary bonds.[1]

About this definition we need for the present only remark that it drops all previous notion, either of "chance," which seems to have dominated his earlier expositions, or of the kindred "randomness" which he seems to have preferred later.[2] And this striking omission of "chance" and "randomness" from his definition would appear to be deliberate. For, in contrast to his previous writings, it holds good throughout his entire present chapter on the topic.

But if this be really so, why then our quarrel dies! Just this property of "chance" and "randomness" has always been that against which our arguments have been directed. And now—much as this new and surprisingly broad definition of "sampling" may dissatisfy us from the standpoint of terminology, it would seem to offer little or no foothold for disputing about the actual facts—or perhaps even for considering them.

Nevertheless, the advocates of the Theory have still some reason to watch their steps. For instance, they seemingly continue to base their theory on the law that the "most probable" frequency of different bonds occurring together equals the product of the most probable frequencies of their occurring respectively at all. In particular, the assumption of this law supplies the whole foundation of their doctrine that different tests have different amounts of "communality,"

[1] See his book, p. 54.

[2] See my account in *Brit. J. Psych.*, 1938, pp. 187–188.

or, as they say, different degrees of "richness."[1] But this law, though true enough of independent and in this sense random occurrences, is quite invalid of non-random ones. Drop the randomness and the bottom of the doctrine falls out.

Again, there is much confusion about the authorship of "group factors." Their introduction has been commonly claimed, both by the proponents of Sampling, as also by the rival advocates of Two Factors. The fact is that the opposing parties have been talking of different things. On the one hand, the group factors fundamental to the sampling theory were first proposed in 1915, and were in every test—as we have seen—prodigiously numerous. Hence the amount of any one of them in a single test could not in general be of appreciable magnitude. Accordingly, not one of them ever has or could be actually measured or even named. On the other hand, the group factors envisaged by the theory of Two Factors have dominated this view ever since its origin in 1904 and have necessarily been measurable every time (or else they would not come into the picture at all) they include such very appreciable influences as those named Speed, Verbality, Space, and Memory. It is only about the group factors of the former kind —microscopic, they may be called—that the notion of randomness has had any plausibility. To explain the other or microscopic kind, mere randomness appears quite inadequate.

But the more we are in this fashion impelled to abandon any attempt to explain mental constitution exclusively by "chance" or "randomness," the more we are obliged to have recourse to its opposite; that is to say, non-randomness, system, structure or organization. This fact not only applies to the whole universe of bonds, neurones, sensations or other suggested ultimate elements of mental activity; it seems to be even more characteristic of the selection made of these elements in the course of constituting either individual minds or mental tests. But, of course, the required organization need not necessarily be that which is involved in the Theory of Two Factors; indeed, it may be of some other kind yet to be conceived.

Accordingly, it is in the positive analysis of complex system, rather than exclusively in such negative and atomistic conceptions as "chance" and "randomness," that would appear to lie the most promising field for further study of ability. And, indeed, at least one initial move in this direction has now been adventured by Thomson himself in his intriguing notion of "sub-pools."

[1] See his book, p. 47.

Be this as it may, however, I would like to reiterate that, after all, the preceding divergencies of my opinion from that of Thomson do not touch any of his admirable work on the main edifice of factorial analysis, as built up by exact observation and sound mathematics. And as for the hypothetical sub-theories which do excite controversy, to have this effect, it seems to me, is only their normal rôle in science. Our arguing is but analogous to that which has lasted for ages about the essential nature of electricity. By the way, we may note here, that now a new and very ingenious hypothesis has been brought into the field by Burt which, he says:

> would seem largely to reconcile Thomson's views with Spearman's.[1]

For yet a further mediation we have to thank the impressive considerations of M. S. Bartlett.[2]

Personally, I suspect that still other theoretical doctrines could be imagined. In fact, I regard them all—even my own!—as no more than temporary theoretical props. Let us by all means ponder and even contend about these speculative matters upon due occasion. But not in such wise, as to impede more positive scientific progress.

[1] *Ibid.*, p. 192.

[2] *Brit. J. Psych.*, 1937, pp. 97–104.

TEACHING PRACTICES

EDWARD LISS

New York City

The analysis of any choice of vocation is subject to many limitations. Not alone is there lack of complete knowledge, which is always the case in any evaluation, but the presence of a complexity of motivations conditions and limits the conclusions, since the whole life experience tempers this choice as it does any other choice of vocation and avocation. None the less, with these knowledgeable limitations a certain degree of clarification is of value.

In the dissection of motivations the exposure of certain segments leads to quite unattractive aspects. To the casual uninitiated observer this is a deterrent, but to the scientific experienced investigator there is a certain beauty in the fusion of these unaesthetic components into a functioning aesthetic entity. With all there is more in this entity than the sum of its parts. No matter how numerous the motivations may be, ego gratification is a dominant and vital urge and its gratification is a fundamental impetus and goal. A successful and healthful emotional outcome is based upon the linking of this inherent need with the gratification which comes through social functioning.

For an adult to function with a certain instrument—in this instance the application of factual material—he must not alone go through a process of accumulation and assimilation, but with it also must go a definite humanizing of that acquisition, in order that he may function successfully with contemporaries and with his students in its application. What has gone before of necessity conditions a favorable or unfavorable outcome. There is a basic constitutional factor, an inherent capacity which to a certain extent is a transmission through physical endowment. There are certain inherent qualities or fundamentals which condition favorably or unfavorably the superimposed structure. Able teachers are artists as well as scientists in whom contemporary practices rise comparatively frictionless from previous growth. In whatever aspect of education these experts may participate, their attitudes towards their colleagues superior in authority or their equals in status play a rôle which determines healthful functioning. Through a process of successful identification with acceptable parental practices to which they have been exposed in times past, they in turn carry on this fundamental human approach in their professional activities.

We are constantly forced by experiences to evaluate the past in each educator's own human relationships so that we may be able to predict healthful practices in their own adult relationships. The attitude of coöperative acceptance of authority and an attitude of social sharing with contemporaries is not a spontaneous acquirement. Early positive healthful acceptance of parental ways helps immeasurably to the acceptance of a rôle of a similar nature with younger people. This privilege of power is a weapon for good or evil, tempered by the healthful or over-compensating urges which every person utilizes in handling other human beings. Where the educator has witnessed in early life abuses of parental privileges in which they themselves have been exploited, they in turn may quite unconsciously take over these techniques in varying degrees, and through them practice a tyranny which may vary from forthrightness to a benevolent despotism. No matter how successfully they may function within the classroom routines, through this preconditioning they are inevitably the victim of their practices, for their own sense of guilt creates not alone emotional tension for others but for themselves. They are prone in their attitudes to be instinctively on the defensive—sometimes belligerently—towards individuals of superior status and they function poorly either because of excessive timidity or over-compensating repressed hostility. This pattern may carry over in their attitudes towards parents, or will bring about negativity and hostility in their human relations with them. With the definite trends now regarded as a part of the educational process the parents have not only a status which must be respected, but also they contribute profoundly to the security of the learning process if there is not a divided philosophy. These elements warrant our specific attention, for more and more we must accept the concept that the acquisition of factual material without its socialization is a profound error in our philosophy of modern education.

In terms of their own early sibling relationships educators may carry over the stability or hostility in their attitudes towards the children. In the ideal state of affairs a transmission from functioning with your own siblings to a successful functioning with other children is a comparatively natural and easy gradation. An early happy period similarly affects the attitude towards students. Where there is unresolved and residual sibling hostility the teacher-pupil relationship is colored by aspects of rivalry and antagonism with their conse-

quent effect on not only social patterns, but with dire effects upon the learning process. Teaching practices, more so than most professional practices which concern themselves with childhood, give the adult an opportunity to function (1) like a good parent; (2) like poor parents with power they cannot control; (3) unlike their own poor parents—a revolt against and a compensation for practices which they feel must not be practiced on other children as it was upon them. Of all fortunate backgrounds the first is outstanding. The third category may lead to all sorts of reformation practices and ultimately to good teaching practices, but usually with stresses and strains which bear heavily upon the adult with this type of motivation. We always come back to the function of a sound early family experience which enables an adult to accept contemporaries coöperatively, and from that acceptance social attitudes flow naturally. Successful experimentation leads to a comparatively frictionless technique. In that relationship to pupils there are many facets, and in a sense it is a childhood renewal. The educator can, although playing a mature rôle, function on that level and still contribute. He can renew and re-experience with his pupils his own childhood, and his ability in this identification fuses effective professional equipment. This ability to revive within himself a child's sense of values enables him to function in depth as well as breadth.

Then again these are the educator's phantom children, a maternal or paternal need which is biologically present in all of us to a varying degree. A primitive biological need, if you will, but there are gradations of this need in all of us, and where it may not be elemental enough for the individual to accept frankly the rôles of parent, it still must be substantial enough in the educator so that teaching relationships serve as a healthful activity. This vicarious parenthood is subject to pressures which to a certain extent are quite insurmountable. Only a realization that the actual parent privileges are not quite transferable leads to a sublimation of the rôle of parent substitute, which is so important if the teacher is to function successfully in humanizing the learning process. Where there are present unresolved conflicts in this field—and there always are—rivalry of various depths predominates in what should be a vital coöperative practice between parent and educator.

Good teachers renew themselves in their pupils. There is the perpetuation of self which students consciously and unconsciously

adopt through identification with good teacher imagos. Just as there is a racial preservation through biological drives, there is a racial preservation of culture through this exposure of child to adult when that exposure is accompanied by healthful relationships. Paradoxically, it is quite possible too that many a poor teacher has aroused a crusading spirit in pupils, which prompts them to function in this field of endeavor so other children will not be exposed to similar unfortunate experiences. This is a negative contribution although it may become a constructive one. The question comes up, in what way does the teacher differ from those in other professions? Primarily there is this fact, that, next to the parent relationship, it is the most intense human contact that growing organisms are exposed to at a time when they are not only most flexible but most impressionable. That this relationship is instrumented by factual material which we call the process of learning, is possibly the only variation from other professional work, and this so-called learning process becomes a sword or a ploughshare instrumented by the energies, the forces which are either tempered by wisdom or sharpened through aggression. Where affection flowers teaching grows. Where there are temperance and tolerance and understanding the learning process expands and ventures into new aspects. Where it is used as an instrument essentially for ego gratification in the teacher, its implications are predominantly hostile, and, no matter what the constructive intent may be, it is devastating in its progress, for the attainment of this goal of self-aggrandizement is at the expense of the student.

Learning is a highly emotionalized procedure and is subject to the laws which govern emotional affects. Its practices may be harnessed to purposeful functioning or they can be an instrument of aggression no less destructive than any other deleterious acquisitive practice. It is the purposeful social application of knowledge, not its acquisition, which determines its value to the individual and particularly to society. Acquisition is only a phase towards purposeful social utilization. Learning can be an egotistical accomplishment or a mature sharing experience. This latter phenomenon is determined by the individual's acceptance of humanizing factors which function in emotional social relationships and which, when incorporated, are conducive to constructive utilization of endowments and acquisitions.

Teaching is a recapitulation of one's own growth, learning a commodity which passes in the exchange between two or more human

beings through which the educator renews his youth and the pupil and educator acquire maturity. It is a highly affective experience and its success is determined by the free flow of impulses which enter into successful functioning, and this free flow is accelerated or retarded by the mature or immature motivation which enters into the human and intellectual relationship of the participants.

BOOK REVIEWS

PAUL MONROE, Editor. *Conference on Examinations.* New York: Bureau of Publications. Teachers College, Columbia University, 1939, pp. 330.

These proceedings of a conference held at Dinard, France, in September, 1938, under the auspices of The Carnegie Corporation, The Carnegie Foundation, and the International Institute of Teachers College suffer from the same weaknesses that limit the usefulness of so many convention reports. There is neither a table of contents nor an index, hence the only way in which a reader can locate what he wants is to turn the pages one by one. The editor was not disposed even to list the titles of the papers read or topics discussed in his outline of the proceedings (p. xi–xiii). A very large part of the text is consumed by conversation and discussion between various members of the conference which makes interesting but not very meaty reading.

The first two sessions were devoted to: (*a*) Consideration of errors made in grading examinations for the baccalaureate in France, (*b*) the application of factorial analysis to pedagogical data, (*c*) the relationship between the École Normale Supérieure and the examination system, and (*d*) the use of tests to direct French children into the proper type of secondary education. The third, fourth and fifth sessions involved a discussion of the research on the validity and reliability of the English essay examinations which had been conducted under the leadership of Sir Philip Hartog and Sir Michael Sadler. The Scottish committee reported upon its study of the intelligence of one thousand children during the sixth session. The discussion centered upon techniques for selecting reliable samples as well as upon the advisability of reporting analytical scores for a complex examination.

The American representatives described the Learned-Wood Pennsylvania College Study, and the so-called Graduate Record Examination during the seventh session. Thorndike's paper, which followed, emphasized his conviction that the most important problem today in the field of mental measurement involves the establishment of equal scale units and the location of a zero point. The discussion was fairly vigorous but involved conventional considerations only.

As was stated above, the brief papers—surprisingly general and nontechnical to be read before a group of experts—and the discussion and chatter that followed make rather interesting reading. Good

stories were reported faithfully and a few penetrating criticisms voiced. It would have been stimulating to have attended the conference. The report of it, however, this reviewer would not elect to read.

STEPHEN M. COREY.
University of Wisconsin.

WILLIS D. ELLIS. *A Source Book of Gestalt Psychology.* New York: Harcourt, Brace and Co., 1938, pp. 403.

"This book is a series of abstracts or summaries of thirty-four articles and one book published in Germany between 1915 and 1929 by the leading exponents of Gestalt psychology and their students." The material is largely experimental. The text is designed to furnish in English a more adequate picture of the transition from the German psychology of 1912 to the present than is available in other sources. The aim to reproduce faithfully the *thought processes* of the authors seems to have been adequately achieved. There is no attempt to evaluate or criticise the material. The book contains three general sections: (1) Theory and physical Gestalten are covered under general problems. (2) Perception, perception of movement, perception and related phenomena, animal experiments, thought, psychical forces and pathological phenomena are considered in the second section. (3) The final part consists of replies to criticisms of the Gestalt approach. Although the original reports are approximately ten times the length of the summaries, the reader seldom senses an undue amount of condensation.

In general, the selection of articles to be included is excellent. Of necessity, a large majority are on perception, the field most readily adapted to Gestalt research. Reports which will be found both interesting and highly stimulating to the psychologist without strong systematic bias include those on transparency, influence of form on the assimulation of color and on brightness contrast, color constancy, figural perceptions in animals, and the effect of certain visual disturbances due to brain injury upon form perception. Only the Gestalt devotee, however, will accept as valid the argument presented in such articles as those on tendencies in figural variation, and in Gestalt factors and repetition.

All readers, irrespective of their systematic viewpoint, will label this treatise as an important job well done. Furthermore, the text is very readable considering the type of material covered. No

psychologist interested in contemporary experimental psychology can afford to dispense with the stimulation that comes with reading this excellent book.

MILES A. TINKER.

University of Minnesota.

J. E. WALLACE WALLIN. *Minor Mental Maladjustments in Normal People.* Durham, N. C.: Duke University Press, 1939, pp. 298.

For many years Wallin has been collecting autobiographical case material from members of his classes in universities in all sections of the country. The majority of his three hundred respondents have been mature students, usually teachers and principals, whose histories are in no way abnormal or pathological. Some of those histories have already been published in the author's *Personality Maladjustments and Mental Hygiene.*

In the present volume about three hundred extracts from the histories are presented under a number of classifications, *e.g.* fears and phobias, anxieties, lack of confidence and bashfulness, feelings of inferiority and inadequacy, obsessions and compulsions, dreams, day-dreaming, emotional disturbances, plus a number of other less frequent conditions. The author has an introductory chapter describing the conditions under which the materials were collected, and discussing the importance of experiential history in the development of behavior. Each chapter has a brief discussion of the problems and the reader is referred to specific pages in a dozen of the most available texts in mental hygiene.

As a supplementary volume in courses in the mental hygiene area, and as a source book of illustrative material this book should certainly prove most useful.

C. M. LOUTTIT.

Indiana University.

I. KIETH TYLER. *Spelling as a Secondary Learning.* New York: Bureau of Publications, Teachers College, Columbia University, 1939, pp. 116.

Defining secondary learning as "those more or less unforeseen changes which take place in boys and girls as a result of school experiences because such learning occurs secondarily to the attainment of other primary objectives" (p. 1) Tyler concludes that sixth-grade social studies taught by the unit method lead to more such incidental

learning of spelling than is the case when the subject-matter is logically organized.

Children and teachers in two Maryland counties were used as subjects during the Spring of 1931. The teachers were assumed to be comparably effective and were admonished not to teach the spelling of any words other than those appearing in a speller used by all pupils. Tyler believes they refrained as directed, but the reviewer is in doubt. In one county the one hundred fifty-one pupils involved were taught logically organized subject-matter, while in the second county the "unit" type of organization was used with six hundred fifty-five pupils. Each type of teaching resulted in a considerable amount of secondary learning of spelling but, as stated above, the unit organization was consistently superior.

The comparability of the difficulty of the words to be spelled in the crucial test for each group was well established considering the inherent difficulty of any such demonstration. With respect to the incidental learning of spelling the brighter students benefited more from the unit method than did the others. The study represents another bit of evidence favoring the superiority of those learning situations which are significant to the learners when compared with those less meaningful.

The chief criticism of Tyler's investigation is the inadequacy of the data advanced to support the contention that the "unit" and "logical" teachers were comparable in general effectiveness. There is some indirect evidence to indicate that such was not the case. For example the "unit" group had but recently coöperated in a county-wide reorganization of the elementary-school curriculum. Furthermore, three-fourths of these "unit" teachers coöperated in the investigation by sending lists of words to the author while less than half of the "logical" group did so. If the "unit" teachers were a more stimulating, progressive crowd this certainly would have had a bearing upon the results regardless of the "unit" *versus* "logical" dichotomy.

STEPHEN M. COREY.

University of Wisconsin.

O. B. DOUGLAS AND B. F. HOLLAND. *Fundamentals of Educational Psychology.* New York; The Macmillan Company, 1938, pp. 598.

"This book is intended for college students who are enrolling for the first course in educational psychology," say the authors in the

preface. If this means students who have had no general psychology then the book is inadequate and such students had better have one or more additional courses. If it means students who have had general psychology then the content is much too elementary and the style too stilted.

There are four parts: (1) The content and scientific basis of educational psychology, (2) environment and heredity (3) learning, and (4) measurement. Reading the book gives one the impression of an extensively classified collection of paragraphs cataloguing many topics of possible educational importance. There is the skeleton of a course but vitality is lacking. The first two parts especially approach an unalphabetized dictionary and not always an accurate one. For example, the cell body of the neurone "corresponds to the nucleus of other types of cells" (p. 77); cortical "neurons appear to be the reacting organs for all conscious states and processes" (p. 90). "An instinct may be described as an inherited *combination* of reflexes which unfold in a given pattern that characterizes a particular species" (p. 117). The fourth part on measurements is also a catalogue in which very few actual tests are described. The best section of the book is the part on learning, in which the discussion covers everything from simple motor coördination through play, emotion, imitation to imaginative learning. In this section, while the organization strikes one as peculiar, there is more vitality and less skeleton.

This reviewer would find it quite impossible to be satisfied with this book as a text.

C. M. Louttit.

Indiana University.